Quantifying sources of variation in multi-model ensembles: A process-based approach

Patrick Denis Sessford

April 2015

Submitted for the degree of Doctor of Philosophy in Mathematics.

Submitted by Patrick Denis Sessford, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics, in April 2015.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

(Signature) .....

Abstract

The representation of physical processes by a climate model depends on its structure, numerical schemes, physical parameterizations and resolution, with initial conditions and future emission scenarios further affecting the output. The extent to which climate models agree is therefore of great interest, often with greater confidence in robust results across models. This has led to climate model output being analysed as ensembles rather than in isolation, and quantifying the sources of variation across these ensembles is the aim of many recent studies. Statistical attempts to do this include the use of variants of the mixed-effects analysis of variance or covariance (mixed-effects ANOVA/ANCOVA). This work usually focuses on identifying variation in a variable of interest that is due to differences in model structure, carbon emissions scenario, etc. Quantifying such variation is important in determining where models agree or disagree, but further statistical approaches can be used to diagnose the reasons behind the agreements and disagreements by representing the physical processes within the climate models. A process-based approach is presented that uses simulation with statistical models to perform a global sensitivity analysis and quantify the sources of variation in multi-model ensembles. This approach is a general framework that can be used with any generalised linear mixed model (GLMM), which makes it applicable to use with statistical models designed to represent (sometimes complex) physical relationships within different climate models. The method decomposes the variation in the response variable into variation due to 1) temporal variation in the driving variables, 2) variation across ensemble members in the distributions of the driving variables, 3) variation across ensemble members in the relationship between the response and the driving variables, and 4) variation unexplained by the driving variables.

The method is used to quantify the extent to which, and diagnose why, precipitation varies across and within the members of two different climate model ensembles on various different spatial and temporal scales. Change in temperature in response to increased CO_2 is related to change in global-mean annual-mean precipitation in a multi-model ensemble of general circulation models (GCMs). A total of 46% of the variation in the change in precipitation in the ensemble is found to be due to the differences between the GCMs, largely because the distribution of the changes in temperature varies greatly across different GCMs. The total variation in the annual-mean change in precipitation that is due to the differences between the GCMs depends on the area over which the precipitation is averaged, and can be as high as 63%. The second climate model ensemble is a perturbed physics ensemble using a regional climate model (RCM). This ensemble is used for three different

applications. Firstly, by using lapse rate, saturation specific humidity and relative humidity as drivers of daily-total summer convective precipitation at the grid-point level over southern Britain, up to 8% of the variation in the convective precipitation is found to be due to the uncertainty in RCM parameters. This is largely because given atmospheric conditions lead to different rates of precipitation in different ensemble members. This could not be detected by analysing only the variation across the ensemble members in mean precipitation rate (precipitation bias). Secondly, summer-total precipitation at the grid-point level over the British Isles is used to show how the values of the RCM parameters can be incorporated into a GLMM to quantify the variation in precipitation due to perturbing each individual RCM parameter. Substantial spatial variation is found in the effect on precipitation of perturbing different RCM parameters. Thirdly, the method is extended to focus on extreme events, and the simulation of extreme winter pentad (five-day mean) precipitation events averaged over the British Isles is found to be robust to the uncertainty in RCM parameters.

Acknowledgements

This thesis is funded by the Natural Environment Research Council, and is a part of the CONVEX research project (www.research.ncl.ac.uk/convex). Thanks to all of the members of the CONVEX project, particularly to Lizzie Kendon and Steven Chan of the UK Met Office, for advice and feedback. Also thanks to Hugo Lambert of Exeter Climate Systems (XCS) for discussion on Chapter 3, and to Joe Osborne for providing and formatting the data used in that chapter.

Many thanks to my supervisors, Chris Ferro and David Stephenson. David's advice and feedback was invaluable when designing applications and presenting results, and Chris' guidance throughout my time at XCS made this thesis possible. Much of the development of the methodology has been driven by Chris' ideas, and our regular meetings were always constructive − thanks for your patience.

I'd also like to thank my parents for letting me move back home to save money while writing up my results after my funding had run out. They haven't charged me rent (yet), but I have helped them move house and dig a pond so I think we're even!

Contents

List of Tables

List of Figures

1 Introduction

1.1 Climate models

Climate models represent the climate of the real world mathematically. Perhaps the simplest climate model balances the energy budget of the planet by using the amount of incoming solar energy and an idealized (zero dimensional) representation of the Earth to determine the global-mean temperature at a given point in time (see Collins, 2007). This requires the albedo and greenhouse effect of the planet to be idealized and estimated. These are then used to determine the amount of incoming solar radiation reflected back into space, the amount absorbed by the planet and emitted back into space as infrared radiation, and the amount retained by the planet. There are, however, spatial and temporal variations driven by complicated feedbacks and interactions between different components of the Earth system and its atmosphere that can only be represented by more sophisticated climate models.

General circulation models (GCMs) are computer models that represent many of the feedbacks and interactions that drive the climate of the real world given different forcings such as increased greenhouse gas concentrations in the atmosphere (see Huebener et al., 2007). Unlike a zero-dimensional energy balance model, GCMs divide the Earth and its atmosphere into a three-dimensional grid. Grid boxes can then differ by various characteristics such as latitude, longitude and orography, and these characteristics will affect the amount of incoming solar radiation that each grid box receives, the air pressure at each grid box, the moisture available, etc. Each grid box is taken as a spatial point, referred to as grid points, and so the resolution depends on the number of grid points that span the Earth and its atmosphere. Large-scale physical processes across space and time, such as the advection of a large mass of air, can be represented across multiple (adjacent) grid points by using differential equations and computing the relevant variables at each time step. Resolving physical processes across multiple grid points is referred to as explicit representation.

GCMs aim to simulate a simplified version of the atmosphere and ocean of the real world. As well as knowledge of the real system, however, GCMs are restricted by computational expense. The finer the resolution (in space and time) of a GCM, the greater the computational cost. Orography is restricted by the spatial resolution of a GCM, and some important physical processes of the climate system (e.g. convection) occur on a scale finer than the resolution of any modern-day GCM (see Stensrud, 2007) and so cannot be resolved explicitly. This has led to the use of regional climate models (RCMs), which can dynamically downscale GCMs by embedding a finer grid

over an area of interest (see Hay and Clark, 2003). In this case the RCM is said to be driven by the boundary conditions of the GCM.

RCMs allow the use of a finer resolution over a sub-global region than GCMs allow, and so can explicitly resolve more physical processes. Despite this, even RCMs are unable to explicitly resolve all important physical processes (e.g. cloud microphysics). For example an RCM spanning Europe may have a spatial resolution of 25 km, which is finer than current GCMs (e.g. see Murphy et al., 2007), but the orographic detail is still restricted and any physical process that occurs on a finer scale cannot be explicitly resolved. Such physical processes include a typical cumulus cloud, which occurs on a scale of less than 1 km (Stensrud, 2007, page 6). The inability of GCMs and RCMs to explicitly resolve some physical processes means that parameterizations are used to represent such subgrid-scale processes (Stensrud, 2007). These parameterizations take the explicitly resolved variables (e.g. temperature and humidity) and use them to simulate the subgrid-scale processes, which can then influence the large-scale variables at the next time step. As discussed later, the parameterization of subgrid-scale processes is a source of much uncertainty in how best to mimic the physics of the real world.

Modern-day GCMs and RCMs can be seen as the accumulation of what is known about the climate system (see Section 2 of Collins, 2007, and references therein for further detail on climate models). It is the simulation of precipitation by these climate models that is the focus of this thesis.

1.2 Why study precipitation in climate models?

The frequency and intensity of precipitation have a great impact on society and the natural world, and can lead to drought at one extreme and flooding at the other. Understanding of the occurrence and severity of rainfall is, therefore, of immense importance. Aside from the loss of life, it is estimated that flooding due to extreme precipitation costs Britain alone one billion pounds on average each year (Hall et al., 2005).

The intensity of precipitation has increased over the second half of the twentieth century in much of the northern hemisphere (Trenberth et al., 2003), including over the United States (Karl et al., 1995; Karl and Knight, 1998) and the United Kingdom (Osborn and Maraun, 2008). Indeed, there is now evidence that anthropogenic climate change increased the risk of extreme precipitation in the UK, such as the extreme precipitation event that caused the floods of October and November 2000 (Pall et al., 2011). Moreover, future precipitation intensity is projected to increase over

the majority of Europe other than in the southern areas of the continent (Beniston et al., 2007; Kendon et al., 2008).

Different seasons are associated with different atmospheric conditions at mid-latitudes, and so trends in precipitation can depend on season. In the UK, winter precipitation has intensified to a greater extent than spring and autumn precipitation over the past hundred years (Osborn and Maraun, 2008). Despite this, the trend in extreme summer rainfall remains unclear, with no significant change detected in observations to date over the UK (Osborn and Maraun, 2008; Fowler et al., 2010), and uncertain modelled projections generally over Europe (Fowler and Ekström, 2009). However in the absence of large-scale atmospheric circulation changes, it is often thought that increases in extreme summer precipitation will occur over northern Scandinavia, with decreases over the Mediterranean, and an increase in the proportion of summer rainfall falling as extreme events over central Europe (Kendon et al., 2009). In winter the pattern for extreme precipitation appears clearer. Increases are projected over most of Europe, with the exception being the south, and these projections are thought to be reliable (Kendon et al., 2008). The projected changes in precipitation are thought to be dominated by a general increase in the atmospheric moisture content as a direct consequence of climate change (Kendon et al., 2009).

With respect to the lack of clarity in projected changes in summer rainfall, a contributing factor is that climate models do not seem to simulate convective precipitation well (Vidale et al., 2003; Déqué et al., 2007; Hohenegger et al., 2008). This leads to a lack of consistency in the projections of future convective precipitation between different models. This is largely because the resolution of climate models is generally too coarse to represent convection explicitly. Convective rainfall is localized and so the cumulus or cumulonimbus clouds form on too fine a scale for climate models to resolve explicitly (also see Spencer and Stensrud, 1998, and Fowler and Ekström, 2009).

Modelled projections of precipitation across the globe have suggested that extreme rainfall will increase in the tropics in the future, along with a decrease at lower quantiles of the frequency distribution of the precipitation, whereas future precipitation is projected to increase over all quantiles at mid- and higher-latitudes (Pall et al., 2007). However, it has been suggested that climate models generally underestimate observed precipitation extremes (Min et al., 2011), and even that simulated extreme precipitation has fundamentally different behaviour than observed. This can be seen because when a generalized extreme value (GEV) statistical model is fitted to simulated extreme precipitation, the GEV model often has a negative value for its

shape parameter, whereas observations lead to a (slightly) positive shape parameter (Fowler et al., 2010). Whether the GEV shape parameter is negative or positive indicates whether or not the severity of extreme events is restricted by the physics of the atmosphere, since a negative shape parameter means that the upper tail is bounded, and so there is a bound on how severe events can be.

The importance of precipitation to society and the natural world, and the observed and projected changes in its characteristics, make understanding the real-world driving physical processes of precipitation important. Aside from making projections, climate models can be used to better understand these real-world processes because variables that are unobserved in the real world can be analysed and related to the simulated precipitation. For example, humidity at different atmospheric levels can be obtained at any grid point across the world in a GCM, as can the precipitation that it simulates. For the output of a climate model to be useful, however, its representation of the physical processes must be realistic enough to relate back to the real world. Therefore understanding to what extent, and why, the uncertainty in the best representation of the physics of the real world by climate models affects their precipitation output is crucial to reduce the uncertainty and make the climate models more realistic.

The rest of this chapter is structured as follows: Section 1.3 briefly outlines why precipitation occurs in the real world. Section 1.4 describes how precipitation is simulated by climate models and which parameters of the models are thought to influence this simulation. This will inform our analysis of climate model parameters later in the thesis. Section 1.5 gives an overview of how climate models differ, and Section 1.6 discusses why quantifying the extent to which these differences between climate models affect their mean precipitation rate (precipitation bias) is useful. Section 1.7 introduces random-effects analysis of variance (random-effects ANOVA), which can be used to quantify the extent to which the differences between climate models affect their precipitation bias. Section 1.8 discusses the limitations of focusing on precipitation bias, and why going beyond the random-effects ANOVA and quantifying the extent to which, and determining why, the differences between climate models affect their representation of driving physical processes of precipitation is crucial in understanding and improving the climate models. Finally, Section 1.9 introduces generalized linear mixed models (GLMMs) and Section 1.10 outlines importance measures. GLMMs and importance measures are tools from the statistical literature and the mathematical modelling literature respectively, and will be used when proposing a general methodology for quantifying the extent to which, and understanding why, the differences between climate models affect their representation of physical processes in the next chapter.

1.3 Precipitation in the real world

Precipitation occurs in the real world when moist air rises such that the moisture condenses around dust and other particles in the atmosphere (called condensation nuclei) into raindrops, forming clouds. The raindrops collide and coalesce until heavy enough to fall to the ground under their own weight. Falling raindrops can further collide and coalesce with each other, and can also be evaporated before reaching the ground. If cold enough, the raindrops can freeze and fall as snow or hail.

As air rises it tends to cool because air pressure is reduced as altitude increases, resulting in the air expanding and reducing its energy per volume adiabatically, which defines its temperature. Therefore for given specific humidity (i.e. a given amount of moisture), as a volume of air rises its relative humidity increases until 100% relative humidity is reached. After this the specific humidity would be greater than the saturation specific humidity (i.e. the moisture holding capacity of the air), and so the moisture condenses.

Using pressure to define humidity mathematically, let $e(T)$ be vapour pressure. Vapour pressure is the pressure exerted by the water vapour molecules, and is a function of temperature (T). Let $e_s(T)$ be saturated vapour pressure (the pressure the water vapour molecules would exert if they saturated the air; also a function of temperature). Relative humidity (RH) is, therefore,

$$RH = \frac{e(T)}{e_s(T)} \times 100 \ ,$$

expressed as a percentage (e.g. Wallace and Hobbs, 2006, page 82). Specific humidity (q) measures the contribution of the vapour pressure to the overall pressure (p) exerted by the air (e.g. see Wallace and Hobbs, 2006, page 80),

$$q = \epsilon \frac{e(T)}{p} \ ,$$

where $\epsilon = 0.62$ is the ratio of the molar masses of dry air and water vapour. Replacing the vapour pressure ($e(T)$) with the saturated vapour pressure ($e_s(T)$) provides the saturation specific humidity, q_s. The capacity of air to hold moisture is defined by the Clausius-Clapeyron relation (see Allen and Ingram, 2002),

$$\frac{\mathrm{d}e_s(T)}{\mathrm{d}T} = \frac{L}{T\Delta v} \ ,$$

where L is the latent heat of vaporization, and Δv is the change in volume of the air between when the water is vapourized and condensed (also see Wallace and Hobbs,

2006, page 98). Assuming that the water vapour behaves like an ideal gas, the saturation specific vapour pressure has an exponential dependence on temperature.

In the atmosphere air can rise because of convection, weather fronts, or the orography of the land. Convective precipitation is generally more localized than precipitation that results from weather fronts or orography, and generally occurs on a fine temporal and spatial scale (see Wallace and Hobbs, 2006, pages 17-18). Air that is warmer, and so less dense, than its surrounding air rises due to convection. This means that if air is in contact with warmer land or sea surface, it conducts heat energy, expands, rises, and is replaced by surrounding cooler air. Convection of moist air can be exaggerated by the latent heat transfer when vapour condenses into liquid, warming the air and causing more convection. This convective process can cause localized extreme rainfall if the vertical temperature gradient through the atmosphere (often referred to as the lapse rate) is strong and there is enough moisture available.

Precipitation that results from ascent of air due to weather fronts or orography is often referred to as dynamical or large-scale precipitation. Large-scale precipitation due to weather fronts occurs when air masses of different temperatures meet. When this occurs the warmer air is forced to rise over the cooler air because the warmer air is less dense. Orographic precipitation occurs when air is advected up high ground. Some studies have attempted to classify observed extreme precipitation events in the UK into convective, frontal, orographic, or a mixture. These can either use subjective judgement (e.g. Hand et al., 2004) or mathematical algorithms (e.g. Little et al., 2008). See Wallace and Hobbs (2006) for further detail on precipitation in the real world and atmospheric physics.

1.4 Simulation of precipitation by climate models

1.4.1 Subgrid-scale parameterization schemes

Depending on the spatial resolution of a climate model, large-scale clouds (i.e. stratus clouds) generally span multiple grid points and their development and evolution can be explicitly resolved, whereas convective clouds (i.e. cumulus clouds) generally occur on subgrid scales and so only their cumulative effects can be resolved implicitly (see Vidale et al., 2003, Déqué et al., 2007, and Hohenegger et al., 2008). The resolution required to explicitly resolve each individual convective updraft is thought by some to be between 25 and 1000 m (Stensrud, 2007, page 188). Regardless of the development of clouds, the microphysics of all clouds must be parameterized in climate models (see Stensrud, 2007, Chapter 8).

It can be assumed that there is no subgrid-scale variability in the cloud microphysical processes, but this assumption is only viable for climate models of resolution finer than 30 km, and even then its appropriateness is questionable (see Stensrud, 2007, pages 264-265). More generally, therefore, climate models account for subgrid-scale variability in cloud microphysics by also parameterizing the proportion of the grid box covered by cloud and scaling the microphysical parameterizations in accordance.

To simulate precipitation, large-scale variables such as relative humidity and saturation specific humidity are used with parameterizations of the various different microphysical processes and interactions that result in the precipitation. These processes include the condensation (see Asai, 1965) of the water vapour in the atmosphere, the collision and coalescence of cloud droplets with each other to form raindrops (called autoconversion) (see Kessler, 1995), the collision and coalescence of raindrops with smaller liquid drops as the raindrops fall (called accretion) (again, see Kessler, 1995), the evaporation of raindrops (see Byers, 1965), and the freezing of water in the atmosphere (see Lin et al., 1983). Observation and mathematical modelling of clouds guide the parameterization of these processes.

In addition to the parameterization of cloud microphysics, convective precipitation is governed by convective parameterization schemes (see Stensrud, 2007, Chapter 6). Indeed, a climate model can recognise when its convective parameterization scheme has caused the ascent that drove the precipitation, and so output from climate models often includes convective precipitation as a separate variable from large-scale precipitation. As a part of a convective parameterization scheme, it is common to compute the convective available potential energy (CAPE) at each atmospheric column of grid points. CAPE is the maximum energy available to an ascending parcel of air. CAPE is computed using simple parcel theory (see Emanuel, 1994) to adiabatically ascend an imaginary parcel of buoyant air from a given starting level to the level at which it is no longer buoyant. Buoyancy is defined by the temperature at which the air parcel would be if it was expanded or compressed adiabatically from its starting pressure and temperature to a given standard temperature (i.e. by its potential temperature) at the different atmospheric levels, and so CAPE is a function of potential temperature at the different levels, height, and acceleration due to gravity (see Stensrud, 2007, pages 188-189). Therefore positive CAPE is required for convection, and convective parameterization schemes can be viewed as removing buoyancy from the atmosphere when the buoyancy occurs. See Chapters 6, 7, and 9 of Stensrud (2007) and references therein for further detail on convective parameterizations, cloud microphysical parameterizations, and cloud cover parameterizations in climate models, and other chapters for information on the parameterization of

other physical processes.

Aside from the parameterizations dealing with convection, cloud microphysics and cloud cover, parameterizations that affect precipitation include that of orographic drag (Chapter 10 of Stensrud, 2007). As moist air is advected up mountains, leading to large-scale precipitation, the friction exerted by the mountain affects the speed at which the air rises. Since the orographic detail of a climate model is restricted by its resolution, this friction must be parameterized. The influence of orographic drag is greater in winter than in summer due to the weaker summer winds, and is negligible when there is little subgrid-scale variation in the height of the terrain. However the drag can be substantially underestimated when the wind is strong or the subgrid-scale variation is large (Stensrud, 2007, page 375).

1.4.2 Subgrid-scale parameterization parameters

With respect to the individual parameters in the parameterization schemes, the entrainment coefficient appears to be influential in how sensitive models are to greenhouse gas concentrations and the overall global temperature (Stainforth et al., 2005; Sanderson et al., 2008b). This is thought to be because the entrainment coefficient controls the profiles of relative humidity within climate models, which are influential in determining the atmospheric responses to greenhouse gas increases (Knight et al., 2007; Sanderson et al., 2008a). However, the entrainment coefficient also has a direct impact on the characteristics of clouds, and this has been found to influence convective rainfall but not particularly large-scale, typically winter, precipitation (Fowler et al., 2010). There is also evidence that the ice-fall speed parameter is important for simulating summer precipitation (Fowler et al., 2010). Increasing the entrainment coefficient appears to result in a heavier extreme (upper) tail to the distribution of simulated precipitation (shown by a greater GEV shape parameter), whereas reducing the ice-fall speed has the same effect.

The entrainment coefficient is a parameter that dictates the behaviour of air immediately outside clouds or air currents, controlling the mixing of the surrounding air, since some of this surrounding air will become part of the cloud or air current (e.g. McGuffie and Henderson-Sellers, 2005). This means that the amount of convective activity is substantially influenced by the entrainment coefficient (Gregory and Rowntree, 1990), and so convective rainfall is affected. The sensitivity of the climate and of convective rainfall to the value of the entrainment coefficient is especially important (and concerning) because its optimal value has high uncertainty even for the current climate (Wilson and Toumi, 2005). Moreover, there is evidence that the parameterization scheme of convective downdrafts is unrealistic in climate models,

and that precipitation can be sensitive to this scheme (Spencer and Stensrud, 1998).

As with the entrainment coefficient, there is evidence that the ice-fall speed parameter influences both climate sensitivity and convective rainfall within climate models (Sanderson et al., 2008b), although perhaps not to the same extent as they are influenced by the entrainment coefficient (Fowler et al., 2010). Reducing the ice-fall speed can result in air remaining moister for longer due in part to increased cloud cover at low levels (Sanderson et al., 2008a), and this can lead to increased precipitation when it falls (Meehl et al., 2005).

Additional noteworthy climate model parameters with respect to precipitation include the accretion coefficient and the scaling factor for emissions from anthropogenic sulphur aerosols. Increasing the accretion coefficient has been found to increase the rate of change of projected future extreme precipitation, whereas increasing the scaling factor for emissions from anthropogenic sulphur aerosols has been found to have the opposite effect, although its impact is less clear (Fowler et al., 2010). One might also expect the number of soil levels accessible for evapotranspiration, the threshold of relative humidity for cloud formation, the flow dependency on this relative humidity threshold, the dependence of stomatal conductance on CO_2, the timescale for CAPE closure, the timescale for dynamical diffusion, the starting level for gravity wave drag, the length of forest roughness, and the surface canopy energy exchange, all to influence the simulation of precipitation in climate models (Elizabeth Kendon, personal communication).

1.5 How climate models differ

There is considerable uncertainty in the values of the parameters of climate models that best mimic the physics of the real world. This uncertainty has led to the use of groups of climate models, referred to as ensembles (e.g. see Collins, 2007, for an overview). Perturbed physics ensembles take a climate model and run it multiple times, each with different parameter values. The values of the parameters are designed to be perturbed in a way that captures some of the uncertainty in the optimal values.

In addition to perturbed physics ensembles, sampling the uncertainty in how best to represent the physics of the real world by climate models can be done with multi-model ensembles. Multi-model ensembles sample the structural uncertainty by including different climate models, which often use different structures of equations and numerical schemes to represent the physics, i.e. separate models differ beyond the values selected for their parameters. The distinction is made here between dif-

ferent climate models and the same model but with perturbed parameter values, although this is not an important distinction for the rationale and methodology of this thesis; it is only important when making statements about the models themselves in the application chapters. Collins (2007) defines parameter uncertainty as "uncertainty in the parameters that control the parameterized physical processes in climate models" and structural uncertainty as "uncertainties in choices made when coding the resolved processes" (i.e. perturbed physics ensembles will share these choices but separate climate models will often not).

Aside from climate model uncertainty, there is uncertainty in the initial state of the real world climate, and uncertainty in future forcings such as greenhouse emissions for when projecting into the future. The former can be sampled by using different initial conditions, and the latter by using different anthropogenic emission pathways (see Moss et al., 2010). Different climate models also operate at different spatial and temporal resolutions. As seen previously, general circulation models can be dynamically downscaled by using models of finer resolution in a reduced area, thereby avoiding the computational cost of running the global model at the fine resolution (e.g. see Wilby and Wigley, 1997).

1.6 Why ask 'to what extent does precipitation bias vary across climate models?'

As a result of the uncertainty in how best to represent the physics of the real world by climate models, different models will simulate different precipitation rates given similar atmospheric conditions, and will simulate these atmospheric conditions with different frequencies. Therefore given a time series from each climate model in an ensemble at a given grid point, quantifying the extent to which the (frequency) distribution of precipitation is represented robustly across the different climate models (i.e. quantifying the ensemble spread) gives a measure of the extent to which the uncertainty in the representation of the physics affects the precipitation that the models simulate (to clarify the terminology further, 'represented robustly across climate models' means 'similar in different climate models'). Focusing on mean precipitation rates, variation across climate models in precipitation bias can be defined as variation in mean precipitation rate due to the differences between the climate models.

The entire range of scientific uncertainty with respect to climate models is not represented by ensembles. Indeed, ensembles are not generally designed to represent all of the uncertainty (Knutti, 2008). Ensembles do however sample some of the uncertainty, and (for example) seasonal forecasts are often more accurate when

ensemble means are taken instead of using the forecasts of single models (e.g. Palmer et al., 2005; and see Tebaldi and Knutti, 2007). This suggests that model errors (partly) cancel with each other.

The extent to which climate model uncertainty should affect confidence in the realism of projections of future climate, when output cannot be validated by observation, is debated (see Räisänen, 2007, Parker, 2011, and Schaller et al., 2011, for discussion). However the uncertainty sampled by ensembles does at least affect the clarity of projections (i.e. the confidence that a projection is not sensitive to model uncertainty). For example if the change in the global-mean annual-mean precipitation in response to a given forcing were to be projected robustly by all climate models, then there would be a consensus on this projection and there can be greater clarity when communicating this to policy makers. Therefore assessing the extent to which precipitation is simulated robustly is important when using climate models to make projections for future climate, as well as to understand current and past climate. On a given spatial scale (e.g. global-mean or a single grid point), quantifying the extent to which precipitation bias varies across different climate models can be done by using the random-effects analysis of variance (random-effects ANOVA) (e.g. see Montgomery, 2009) class of statistical models.

1.7 Random-effects ANOVA

The random-effects ANOVA class of statistical models can be used to quantify the extent to which precipitation bias varies across the members of a climate model ensemble by treating each ensemble member as a level of a random (categorical) factor (see Gelman, 2005, for an overview of random factors). A climate model ensemble is not an exhaustive collection of all possible models, and interest is not generally restricted to only the ensemble members available. In other words, a climate model ensemble can be thought of as one possible collection of models from a wider population of all possible ensemble members, and the interest is often in this wider population. Therefore an ensemble lends itself to being statistically modelled as a random factor, e.g. using a random-effects ANOVA. It may be that assuming that the ensemble members are independent random selections from the wider population is not correct, perhaps due to structural similarities between some ensemble members and not others (see Tebaldi and Knutti, 2007; Knutti, 2010; Masson and Knutti, 2011), but the random-effects ANOVA can still be used to provide an approximation to the true process if the fit of the ANOVA to data is of sufficient quality. The random-effects ANOVA allows inference to be made about the wider population of all possible climate models that could be designed and used in the ensemble. Moreover the value of a random effect can be estimated for each individual

ensemble member (see Gelman, 2005), and so by analysing these point estimates (instead of the estimated variance component of the random effect) the focus can be on only the specific ensemble members if required. Many recent studies, including Sain and Furrer (2010), Christensen and Sain (2011), Geinitz et al. (2012) and Furrer et al. (2012), have used random effects to treat climate models as random selections from a wider population.

Let Y_{ij} be a random variable denoting the precipitation on a given spatial and temporal scale at time i in climate model j. For example Y_{ij} could be the global-mean annual-mean precipitation in year i in ensemble member j, and a time series from each ensemble member could be used to estimate its expectation (μ_j) in member j and the internal variation (τ_ϵ) within each ensemble member. The one-way random-effects ANOVA class of statistical models can then be defined as

$$\begin{aligned} Y_{ij} \mid \psi_j^{(0)} &\sim (\mu_j, \tau_\epsilon) \\ \mu_j &= \beta_0 + \psi_j^{(0)} \\ \psi_j^{(0)} &\sim (0, \tau_0) \ . \end{aligned}$$

In this model the conditional distribution of Y_{ij} given $\psi_j^{(0)}$ has mean μ_j and variance τ_ϵ, where μ_j is composed of a fixed effect, β_0, and $\psi_j^{(0)}$, which is a random effect with mean 0 and variance τ_0. The variation (τ_0) in $\psi_j^{(0)}$ estimates the variation in mean precipitation in the wider population from which the ensemble members came. As such, τ_0 accounts for different climate models that could have been designed and used in the ensemble. For example in a multi-model ensemble different parameterization schemes could have been used, or in a perturbed physics ensemble the physical parameters could be have been perturbed differently.

Assuming the one-way random-effects ANOVA, there are two sources of variation that contribute to the total variation in the precipitation: the variation across the different climate models in precipitation bias (quantified by τ_0), and the internal variation within each climate model (quantified by τ_ϵ). The internal variation within each climate model can also be viewed as the residual variation from the ANOVA, i.e. the variation in precipitation that is unaccounted for by the variation across the climate models in the precipitation bias. The relative contributions of the two sources of variation to the total variation in the precipitation are (e.g. Cox and Solomon, 2003)

$$\rho_\psi = \frac{\tau_0}{\tau_0 + \tau_\epsilon} \qquad \text{and} \qquad \rho_\epsilon = \frac{\tau_\epsilon}{\tau_0 + \tau_\epsilon} \ ,$$

where ρ_ψ is the contribution of the variation across climate models in precipitation bias, and ρ_ϵ is the contribution of the internal variation within each climate model. The proportion of the total variation in precipitation due to the variation across climate models in precipitation bias is, therefore, quantified by ρ_ψ. This provides an estimate for the extent to which the (frequency) distribution of precipitation is simulated robustly across climate models.

1.8 Why ask 'to what extent, and why, does the representation of driving physical processes of precipitation vary across climate models?'

Quantifying the extent to which precipitation bias varies across climate models provides insight into how robustly precipitation is simulated across the climate models. However there is more to the simulation of precipitation than the bias in its distribution. It could be that the physical processes that drive precipitation are not represented robustly across climate models, but that this does not show up in the mean over a given time period because it is compensated for by additional bias. In this scenario, the precipitation cannot be said to be simulated robustly across the climate models in any general sense. It is robustness of the representation by climate models of the driving physical processes that will lead to different climate models simulating similar precipitation totals in given atmospheric conditions. It is also the representation of physical processes that drive changes in precipitation in climate models in future climates. Therefore analysing these processes within climate models will help to determine why the precipitation varies across the models and how the uncertainty in how to best mimic the physics of the real world influences the precipitation.

Statistical models can be used to capture the representation of physical processes within climate models by building on the random-effects ANOVA discussed in Section 1.7 and using predictor variables of precipitation. More precisely, given a time series from each climate model in an ensemble, the temporal variation in the predictors can be used to capture temporal variation in the precipitation. Additionally, multiple random effect terms can be used to allow variation across ensemble members in the representation of the physical process due to the differences between the ensemble members. Using predictors and random effects in statistical models in this way can be done with the generalized linear mixed model (GLMM) (see Breslow and Clayton, 1993) class of statistical models. By capturing a driving physical process of precipitation within climate models using a GLMM, the extent to which the representation of the physical process varies across climate models can be quanti-

fied. In doing this, the robustness of the simulation of precipitation across climate models can be rigorously assessed, and insight into why precipitation varies across climate models can be gained. Moreover the influence of each predictor can be assessed to help understand the temporal variation and identify the dominant drivers of precipitation.

1.9 GLMMs

Random-effects ANOVAs work by comparing their variance components (τ_0 and τ_ϵ in Section 1.7), which are the only sources of variation present in the statistical model. When using predictors and random effects to capture the representation of a physical process by climate models with a GLMM, however, there are multiple sources of variation present that are not immediately comparable. These sources of variation are 1) the temporal variation in each predictor, 2) the variation across climate models in the distribution of each predictor, 3) the variation across climate models in the relationship between each predictor and the precipitation, 4) further variation across climate models in precipitation bias, and 5) the further internal variation in the precipitation that is unaccounted for by the physical process or precipitation bias. Each of these sources of variation contributes to the total variation in the precipitation. It is sources (2) to (4) that capture the variation in how different climate models represent the physical process (i.e. that due to climate model uncertainty), but source (1) also provides insight into how the climate models represent the temporal variation in the physical process and can be thought of as variation within each ensemble member in the conditions of the atmosphere (i.e. the atmospheric conditions that could be varied when different initial conditions are set in climate models, known as internal variability uncertainty; see Yip et al., 2011, and references therein). Therefore the goal is to extend the random-effects ANOVA to compare the sources (1) to (5), combinations of them, and interactions between them. More specifically, it is not the absolute size of each source of variation that is of interest here, it is their effect on the simulation of precipitation.

Sources of variation (1) and (2) above are both from the variation in the predictors used by a GLMM, i.e. the effects of both sources have to be inferred from the variation in each predictor. However making the distinction between these two sources is important when quantifying how robustly a physical process is represented across different climate models because whereas source (1) is temporal variation within each climate model, source (2) is variation due to the differences between climate models.

See Figure 1.1 for a schematic of the sources of variation captured by a GLMM with

one predictor and only two climate models in the ensemble. In this schematic there is variation (represented by γ, which varies horizontally) across the two climate models in the distribution of predictor X, further (temporal) variation in X, variation (represented by $\psi^{(1)}$, which captures variation in the slope) across the climate models in the relationship between X and Y, further variation (represented by $\psi^{(0)}$, which varies vertically to capture variation in the intercept) across the climate models in bias, and additional (unaccounted for) variation in response Y. Each source contributes to the total variation in Y, which is represented by the probability density function (PDF) of Y shown (in blue) vertically to the right of the schematic. For example suppose that there was no variation across the two climate models in the relationship between X and Y (i.e. $\text{var}(\psi^{(1)}) = 0$), such that the slope of the red climate model was reduced to that of the green model. In this scenario there would be less total variation in Y because its greatest values would be reduced towards its mean.

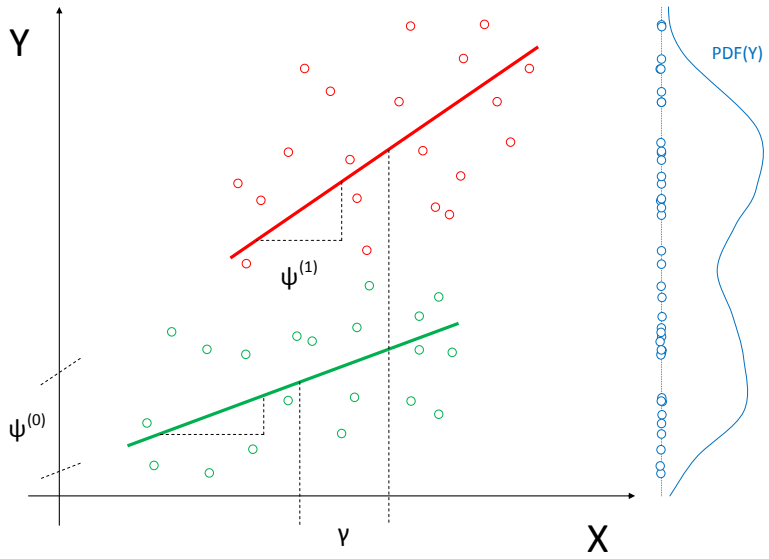

Figure 1.1: Schematic showing the sources of variation in Y captured by a GLMM fitted to a two-member (shown in green and red) ensemble of climate models using a single predictor (X). See text for details.

Whereas the influence on precipitation of the two sources of variation present in the random-effects ANOVA shown in Section 1.7 can be compared by directly comparing the two variance components, there is no immediate way to compare the influence on precipitation of each of the sources of variation present in a GLMM. To see this, consider a GLMM that relates two predictors ($X_{ij}^{(1)}$ and $X_{ij}^{(2)}$) to precipitation (Y_{ij}) on a given time scale (e.g. daily totals) at a given grid point, where i indexes time and j indexes climate model. Suppose that the specification of this GLMM is

$$
\begin{aligned}
Y_{ij} \mid \left(\psi_j^{(0)}, \psi_j^{(1)}, \psi_j^{(2)}, X_{ij}^{(1)}, X_{ij}^{(2)}\right) &\sim \text{Gamma}(\mu_{ij}, \lambda) \\
\text{where } \log(\mu_{ij}) &= \beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \psi_j^{(0)} + \psi_j^{(1)} X_{ij}^{(1)} + \psi_j^{(2)} X_{ij}^{(2)} \ , \\
\left(\psi_j^{(0)}, \psi_j^{(1)}, \psi_j^{(2)}\right) &\sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}) \ ,
\end{aligned}
$$

μ_{ij} denotes the fitted expectation of the precipitation, λ denotes the dispersion parameter of each fitted Gamma distribution, and $\boldsymbol{\Sigma}$ denotes the (multivariate) variance components of the random effects $\psi_j^{(0)}$, $\psi_j^{(1)}$ and $\psi_j^{(2)}$. With this GLMM specification there is no simple way to read off the size of the influence on precipitation of the variation in $X_{ij}^{(1)}$ compared to the influence of the variation in the relationship between $X_{ij}^{(2)}$ and the precipitation (captured by $\psi_j^{(2)}$), for example, because the variance component of $\psi_j^{(2)}$ is on a scale that is incomparable to the coefficient (β_1) of $X_{ij}^{(1)}$, or the variation in $X_{ij}^{(1)}$. Indeed, the influence on precipitation of many of the sources of variation present in the GLMM cannot be compared simply. Moreover, the GLMM specification assumes linearity between each predictor and the linear predictor ($\log(\mu_{ij})$), and assumes a constant dispersion parameter (λ). If these assumptions are relaxed by adding further terms to the specification then this could only complicate comparisons between the sources of variation further.

1.9.1 Comparing sources of variation in a GLMM

To compare the size of different sources of variation in statistical models, the random-effects ANOVA has been extended to include a variety of more complicated, but related, statistical models. Gelman (2005) outlines the use of a random-effects ANOVA when multiple random effect terms are present, although does not generalize to time series where random effects are used to allow coefficients of predictors to vary. Geinitz et al. (2012) built on Gelman (2005) by extending the approach to the multivariate (i.e. multiple response variables to the statistical model) case. Other variants of the random-effects ANOVA include the functional ANOVA (Ramsay and Silverman, 2005, Chapter 13), where functional responses allow an analysis of variance in different circumstances (e.g. at different spatial locations, or different times). Also see Kaufman and Sain (2010), who developed the functional ANOVA, and Sain et al. (2011) for the use of a functional ANOVA with climate model data. The focus, however, of these extensions to the random-effects ANOVA is on using multiple random effects that allow variation across different random factors (rather than a single random factor, i.e. the climate model, indexing multiple random effects), and on multivariate or functional response variables, but not on GLMMs.

A difference between the variants of the random-effects ANOVA discussed above and GLMMs is the presence of predictors. Predictors are generally on different scales and units, but their influence on the response can often be compared by standardizing them, transforming their coefficients (e.g. see DeMaris, 1993), or by standardizing their coefficients (e.g. see Gelman, 2008). However, these methods are designed to make the effects of the predictors comparable with each other, not with other sources of variation in the statistical model, i.e. not with the random effects or residual variation. Gelman and Pardoe (2007) discuss limitations of transformed and standardized coefficients, as well as partial derivatives (e.g. see Roneck, 1991). Without the ability to compare all sources of variation in a GLMM, the extent to which the process captured by the GLMM is represented robustly across climate models cannot be quantified.

Referring to sources of variation as inputs, the difficulty of quantifying the effect of inputs on the response variable in some classes of statistical models (including GLMMs) motivated Gelman and Pardoe (2007) to propose what they call average predictive comparisons. These measures are based on the expected difference in the response variable of a statistical model from perturbing each input, averaged over an assumed (population) distribution for the inputs. Average predictive comparisons allow the comparison of the influence on the response of each input, although the authors do not consider the measures for multiple inputs simultaneously. Average predictive comparisons aim to quantify the expected change in the response variable from perturbing each input, and as such do not aim to quantify the contribution to the total variation in the response variable of the variation in each input. Therefore average predictive comparisons are not designed to be the extension to the random-effects ANOVA that we require.

A method that could be used to quantify the impact of each input on the response variable, and could be seen as an extension to the random-effects ANOVA, is the use of analysis of deviance tables (see Hastie and Pregibon, 1992, Chapter 6). Analysis of deviance tables are similar to analysis of variance tables (see Chambers, 1992, Chapter 4), but can be used with a wider class of statistical models, including GLMMs. Whereas analysis of variance tables are based on comparing the change in residual sum of squares when inputs are added or removed from the statistical model (i.e. comparing nested statistical models), analysis of deviance tables compare the change in deviance. The deviance (D) for empirical data y of a model with parameters $\hat{\theta}_0$ is

$$D(y) = -2\left\{\log[p(y \mid \hat{\theta}_0)] - \log[p(y \mid \hat{\theta}_s)]\right\} ,$$

where p denotes probability density and $\hat{\theta}_s$ denotes the parameters of the saturated

model (i.e. with a parameter for every empirical datum). Deviance is based on the probability of the empirical data, as estimated by the fitted model. As such, the deviance of a model is affected by the appropriateness of the assumptions on which it is based. As a result, adding or removing inputs will change both how close the fitted values are to the empirical data, and will change the appropriateness of the assumptions that underpin the GLMM. Therefore, analysis of deviance tables are useful for deciding on which inputs to include in the GLMM, but their use in quantifying the proportion of variation in the response variable accounted for by the variation in different inputs is limited.

Statistically rigorous methods to quantify the contribution of the variation in each input to the total variation in the output of a mathematical model have, however, already been developed under the banner of sensitivity analyses (see Saltelli et al., 2000). These methods are based on simulation from a single mathematical model and use terms called importance measures, which are discussed in the next section. In the literature covered in the discussion below, the importance measures tend to be used with deterministic mathematical models (i.e. with no residual term) rather than statistical models. Despite this, they generalize to any mathematical model and so can be used with GLMMs by treating the residual term as an input. Importance measures partition the total variation in the response variable of a GLMM in a way that provides the same result as a random-effects ANOVA in the two-input case shown in Section 1.7. We will outline this in Chapter 2, and as a result importance measures will form the basis of the methodology that we propose for analysing how robustly physical processes are represented across climate models, and why they differ.

1.10 Importance measures

1.10.1 Defining importance measures

Given a mathematical model relating inputs to an output variable of interest, sensitivity analyses can quantify the influence of the variation in each input on the output. To achieve statistical rigour, however, the sensitivity analysis must be global, not local (Saltelli and Annoni, 2010). Local sensitivity analyses hold each input to a nominal value, perhaps its mean, then vary one input at a time and observe the resulting effect on the output of the model (see Murphy et al., 2004, or Hof et al., 2008, for examples of uses of local sensitivity analyses). This is generally not statistically rigorous because interaction between inputs is unaccounted for. Global sensitivity analyses are also based on varying different inputs and observing the effect on the

output, but are more rigorous because they take into account the whole distribution of all the inputs. Global sensitivity analyses are based on the law of total variance (e.g. Saltelli et al., 2010), which is a mathematical law that relates two variables to one another (e.g. Weiss, 2005). Simulation can be used to explore the input space to the mathematical model thoroughly (see Jansen, 1999).

The law of total variance states that the total variation in the output of a mathematical model can be orthogonally decomposed into the variation attributable to an input and the remaining variation in the output. This is achieved by conditioning on the input and computing the variation in the conditional expectation of the output, and the expectation of the conditional variation in the output, respectively. To demonstrate this, let Y denote the output of a mathematical model and X one of its inputs, then

$$\text{var}(Y) = \text{var}_X[E_Y(Y \mid X)] + E_X[\text{var}_Y(Y \mid X)] \,. \tag{1.1}$$

In Equation (1.1), $\text{var}(Y)$ is the total variation in the output of the mathematical model, $\text{var}_X[E_Y(Y \mid X)]$ is the variation in Y accounted for by the variation in input X, and $E_X[\text{var}_Y(Y \mid X)]$ is the remaining variation in Y (e.g. Weiss, 2005). This can be seen because if X accounts for much of the variation in Y, then the value expected for Y will vary greatly depending on the value of X (and so $\text{var}_X[E_Y(Y \mid X)]$ will be high). Conversely, if X accounts for little of the variation in Y, then $E_Y(Y \mid X)$ will simply be close to $E(Y)$ irrespective of the value of X (and so $\text{var}_X[E_Y(Y \mid X)]$ will be low). The subscripts used here emphasize that the expectation is for Y in $E_Y(Y \mid X)$, but that the variation in this is then taken over different values of X. If $\text{var}_X[E_Y(Y \mid X)]$ is the variation in Y accounted for by the variation in X, then $E_X[\text{var}_Y(Y \mid X)]$ is the variation in Y that can be accounted for by the variation in the other inputs.

Viewing $\text{var}_X[E_Y(Y \mid X)]$ and $E_X[\text{var}_Y(Y \mid X)]$ as above highlights that $\text{var}_X[E_Y(Y \mid X)]$ is a summary measure that can quantify the extent to which the variation in X influences Y. Moreover, however, terms of the form $E_{\cdot}[\text{var}_Y(Y \mid \cdot)]$ are also useful. Using $-X$ to denote all of the inputs except for X, a different decomposition to that of Equation (1.1) could be

$$\text{var}(Y) = \text{var}_{-X}[E_Y(Y \mid -X)] + E_{-X}[\text{var}_Y(Y \mid -X)] \,. \tag{1.2}$$

In Equation (1.2), $\text{var}_{-X}[E_Y(Y \mid -X)]$ is the variation in Y accounted for by the variation in all of the inputs $(-X)$ except for X. As such, the remaining variation, $E_{-X}[\text{var}_Y(Y \mid -X)]$, can be seen as the variation in Y accounted for by the variation in X. Therefore we now have two measures that summarize the extent to which the

variation in X influences Y, $\text{var}_X[E_Y(Y \mid X)]$ and $E_{-X}[\text{var}_Y(Y \mid -X)]$.

The measure of the extent to which the variation in X influences Y provided by Equation (1.1) above, $\text{var}_X[E_Y(Y \mid X)]$, is called the first-order effect of X (e.g. see Jansen, 1999, and Chan et al., 2000). Following from its mathematical definition, a formal interpretation for the first-order effect of X is

Interpretation 1.1. the reduction in variation in Y expected by fixing X constant.

To see this, the law of total variance (Equation (1.1)) can be rearranged to make the first-order effect the subject,

$$\text{var}_X[E_Y(Y \mid X)] = \text{var}(Y) - E_X[\text{var}_Y(Y \mid X)] \,,$$

where the second term on the right-hand side is the expected variance in Y remaining after fixing X.

The measure of the extent to which the variation in X influences Y provided by Equation (1.2) above, $E_{-X}[\text{var}_Y(Y \mid -X)]$, is called the total effect of X (e.g. see Jansen, 1999, and Homma and Saltelli, 1996). A formal interpretation of the total effect of X is

Interpretation 1.2. the reduction in the variation in Y from averaging over X.

As with the first-order effect above, to see this rearrange the law of total variance (this time Equation (1.2)) to make the total effect the subject,

$$E_{-X}[\text{var}_Y(Y \mid -X)] = \text{var}(Y) - \text{var}_{-X}[E_Y(Y \mid -X)] \,,$$

where the second term on the right-hand side is the variance in Y remaining after averaging over X.

First-order effects and total effects are referred to as importance measures (see Homma and Saltelli, 1996). The difference between the first-order effect and total effect of X is that its total effect accounts for the variation in Y due to any interactions between X and the other inputs, and its first-order effect does not (Jansen et al., 1994; Homma and Saltelli, 1996). We mathematically define interactions below, but interaction between inputs of a mathematical model occurs when the inputs share a multiplicative term in the model (which could be due to a non-linear function within the model) or any other non-additive combination between inputs (e.g. an input raised to the power of another). For example the inputs ($X^{(1)}$ and $X^{(2)}$) interact in the mathematical model $Y = X^{(1)}X^{(2)}$ because the inputs are multiplied by each other.

The total variation in the output Y of a mathematical model decomposes into the first-order effect of each of its inputs ($X^{(i)}$, $i = 1, ..., k$) and any interactions between these inputs. This decomposition is valid only if we assume that the inputs are independent. To see this, to define the interactions, and to show that interactions are included in total effects, we now summarize the decomposition equations of Sobol' (1993) (e.g. Section 2.1 of Homma and Saltelli, 1996, Section 2 of Jacques et al., 2006, and Section 2 of Saltelli et al., (2010), each provide similar summaries of the decomposition equations). Output $Y = f(X^{(1)}, ..., X^{(k)})$ is a function of its k inputs, and Sobol' (1993) showed that Y can be decomposed into the sum of functions of the inputs. These functions can be arranged into summands of increasing dimensionality, such that

$$\begin{aligned} Y &= f(X^{(1)}, ..., X^{(k)}) \\ &= f_0 + \sum_i f_i(X^{(i)}) + \sum_i \sum_{j>i} f_{ij}(X^{(i)}, X^{(j)}) + \cdots + f_{1...k}(X^{(1)}, ..., X^{(k)}) \ . \end{aligned}$$

There are a total of 2^k terms in this decomposition, $2^k - 1$ functions of the inputs and f_0. These terms are

$$\begin{aligned} f_0 &= E(Y) \\ f_i(X^{(i)}) &= E(Y \mid X^{(i)}) - E(Y) \\ f_{ij}(X^{(i)}, X^{(j)}) &= E(Y \mid X^{(i)}, X^{(j)}) - f_i(X^{(i)}) - f_j(X^{(j)}) - E(Y) \\ &= E(Y \mid X^{(i)}, X^{(j)}) - E(Y \mid X^{(i)}) - E(Y \mid X^{(j)}) + E(Y) \ , \end{aligned} \tag{1.3}$$

and similarly for terms involving more than two inputs. Each of the functions is orthogonal with the others, and the expectation of each is zero. To see this, for example, consider the expectation of $f_i(X^{(i)})$,

$$\begin{aligned} E[f_i(X^{(i)})] &= E_{X^{(i)}}[E(Y \mid X^{(i)}) - E(Y)] \\ &= E_{X^{(i)}}[E(Y \mid X^{(i)})] - E(Y) \\ &= E(Y) - E(Y) \\ &= 0 \ , \end{aligned}$$

and similarly for the other functions. Therefore given that the inputs are independent,

$$E[f_i(X^{(i)}) f_j(X^{(j)})] = E[f_i(X^{(i)})] E[f_j(X^{(j)})] = 0 \ .$$

The first-order effect of an input is the variation in the function involving only that input, and an interaction between given inputs is the variation in the function that

includes only those inputs. Therefore the first-order effect of input $X^{(i)}$ is

$$\text{var}[f_i(X^{(i)})] = \text{var}_{X^{(i)}}[E_Y(Y \mid X^{(i)})] \,,$$

and the interaction between inputs $X^{(i)}$ and $X^{(j)}$ is defined to be (from Equation (1.3) above, excluding subscripts)

$$\begin{aligned}
&\text{var}[f_{ij}(X^{(i)}, X^{(j)})] \\
&= \text{var}\left[E(Y \mid X^{(i)}, X^{(j)}) - E(Y \mid X^{(i)}) - E(Y \mid X^{(j)}) + E(Y)\right] \\
&= \text{var}[E(Y \mid X^{(i)}, X^{(j)})] + \text{var}[E(Y \mid X^{(i)})] + \text{var}[E(Y \mid X^{(j)})] + \text{var}[E(Y)] \\
&\quad - 2\text{cov}[E(Y \mid X^{(i)}, X^{(j)}), E(Y \mid X^{(i)})] - 2\text{cov}[E(Y \mid X^{(i)}, X^{(j)}), E(Y \mid X^{(j)})] \\
&\quad - 2\text{cov}[E(Y \mid X^{(i)}, X^{(j)}), E(Y)] - 2\text{cov}[E(Y \mid X^{(i)}), E(Y \mid X^{(j)})] \\
&\quad - 2\text{cov}[E(Y \mid X^{(i)}), E(Y)] - 2\text{cov}[E(Y \mid X^{(j)}), E(Y)] \\
&= \text{var}[E(Y \mid X^{(i)}, X^{(j)})] + \text{var}[E(Y \mid X^{(i)})] + \text{var}[E(Y \mid X^{(j)})] \\
&\quad - 2\text{cov}[E(Y \mid X^{(i)}, X^{(j)}), E(Y \mid X^{(i)})] - 2\text{cov}[E(Y \mid X^{(i)}, X^{(j)}), E(Y \mid X^{(j)})] \,,
\end{aligned} \tag{1.4}$$

by orthogonality, where

$$\begin{aligned}
&\text{cov}[E(Y \mid X^{(i)}, X^{(j)}), E(Y \mid X^{(i)})] \\
&= E[E(Y \mid X^{(i)}, X^{(j)})E(Y \mid X^{(i)})] - E[E(Y \mid X^{(i)}, X^{(j)})]E[E(Y \mid X^{(i)})] \\
&= E[E(Y \mid X^{(i)})^2] - E[E(Y \mid X^{(i)})]^2 \\
&= \text{var}[E(Y \mid X^{(i)})] \,,
\end{aligned}$$

and similarly for $\text{cov}[E(Y \mid X^{(i)}, X^{(j)}), E(Y \mid X^{(j)})]$. Therefore Equation (1.4) becomes

$$\begin{aligned}
&\text{var}[f_{ij}(X^{(i)}, X^{(j)})] \\
&= \text{var}_{X^{(i)},X^{(j)}}[E_Y(Y \mid X^{(i)}, X^{(j)})] - \text{var}_{X^{(i)}}[E_Y(Y \mid X^{(i)})] - \text{var}_{X^{(j)}}[E_Y(Y \mid X^{(j)})] \,,
\end{aligned} \tag{1.5}$$

and interactions involving more than two inputs are similarly defined. The variation in each of the functions sum to give the total variation in output Y,

$$\begin{aligned}
&\text{var}(Y) \\
&= \text{var}[f(X^{(1)}, ..., X^{(k)})] \\
&= \sum_i \text{var}\Big[f_i(X^{(i)})\Big] + \sum_i \sum_{j>i} \text{var}\Big[f_{ij}(X^{(i)}, X^{(j)})\Big] + \cdots + \text{var}\Big[f_{1\ldots k}(X^{(1)}, ..., X^{(k)})\Big] \,.
\end{aligned} \tag{1.6}$$

Recalling Equations (1.1) and (1.2), only X was conditioned on in Equation (1.1) and all inputs except for X were conditioned on in Equation (1.2). However a strength of using importance measures is that any number or combination of inputs can be conditioned on simultaneously (see Saltelli et al., 2010). Equation (1.5) above shows that conditioning on two inputs ($X^{(1)}$ and $X^{(2)}$) simultaneously, their first-order effect ($\text{var}_{X^{(1)},X^{(2)}}[E_Y(Y \mid X^{(1)}, X^{(2)})]$) consists of the variation in Y due to 1) the variation in $X^{(1)}$ alone, 2) the variation in $X^{(2)}$ alone, and 3) any interaction between $X^{(1)}$ and $X^{(2)}$. Intuitively, the first-order effect ($\text{var}_{X^{(1)},X^{(2)}}[E_Y(Y \mid X^{(1)}, X^{(2)})]$) of $X^{(1)}$ and $X^{(2)}$ can be seen to include the variation in Y due to any interaction between $X^{(1)}$ and $X^{(2)}$ because the variation in $X^{(1)}$ and that in $X^{(2)}$ is averaged over simultaneously, not separately. This means that the joint variation in $X^{(1)}$ and $X^{(2)}$ contributes to the variation taken in Y. Equation (1.5) also highlights that the contribution to the total variation in Y of an interaction between inputs is non-negative because the contribution is the variation in a function. Additionally, it is clear that the first-order effect of $X^{(1)}$ and $X^{(2)}$ decomposes into the sum of the individual first-order effects of the two inputs if $X^{(1)}$ and $X^{(2)}$ do not interact.

As seen previously in Equation (1.2), it follows from the law of total variance that the first-order effect ($\text{var}_{-X}[E_Y(Y \mid -X)]$) of all inputs except for X and the total effect ($E_{-X}[\text{var}_Y(Y \mid -X)]$) of X sum to the total variation in Y. Therefore if $\text{var}_{-X}[E_Y(Y \mid -X)]$ is the sum of each of the individual first-order effects of inputs $-X$ and interactions between them, then the total effect ($E_{-X}[\text{var}_Y(Y \mid -X)]$) of X is the sum of its first-order effect and any interaction that involves X. This means that, for example, the total effect ($E_{-X^{(1)},-X^{(2)}}[\text{var}_Y(Y \mid -X^{(1)}, -X^{(2)})]$) of $X^{(1)}$ and $X^{(2)}$ consists of the variation in Y due to 1) the variation in $X^{(1)}$ alone, 2) the variation in $X^{(2)}$ alone, 3) any interaction between $X^{(1)}$ and $X^{(2)}$, and 4) any further interactions involving $X^{(1)}$ or $X^{(2)}$ with other inputs. This can be seen using a simple example and Equation (1.6). Suppose $Y = X^{(1)}(X^{(2)} + X^{(3)})$, then

$$\begin{aligned}\text{var}(Y) = \text{var}[f_1(X^{(1)})] + \text{var}[f_2(X^{(2)})] + \text{var}[f_3(X^{(3)})] + \text{var}[f_{12}(X^{(1)}, X^{(2)})] \\ + \text{var}[f_{13}(X^{(1)}, X^{(3)})]\end{aligned}$$

because the only interactions between the three inputs are that between $X^{(1)}$ and $X^{(2)}$, and that between $X^{(1)}$ and $X^{(3)}$. The total effect of $X^{(1)}$ and $X^{(2)}$ and the first-order effect of $X^{(3)}$ sum to the total variation in output Y, by the law of total variance. As seen above, the first-order effect of $X^{(3)}$ only accounts for $\text{var}[f_3(X^{(3)})]$. This means that $\text{var}[f_1(X^{(1)})]$, $\text{var}[f_2(X^{(2)})]$, $\text{var}[f_{12}(X^{(1)}, X^{(2)})]$ and $\text{var}[f_{13}(X^{(1)}, X^{(3)})]$ are all accounted for by the total effect of $X^{(1)}$ and $X^{(2)}$.

It follows from the Sobol' decomposition equations that the first-order effect of an

input X can usefully be interpreted as the variation in Y *solely* due to the variation in X (i.e. due to the variation in X alone, e.g. see Homma and Saltelli, 1996). Intuitively, the first-order effect ($\text{var}_X[E_Y(Y \mid X)]$) of a single input X does not include the interactions between X and the other inputs ($-X$) because the variation in Y due to that in X is only evaluated after averaging over the variation in $-X$. The result is that the joint variation in X and $-X$ does not contribute to the variation taken in Y. By contrast, the total effect ($E_{-X}[\text{var}_Y(Y \mid -X)]$) of X takes the variation in Y due to that in X without having averaged over the variation in $-X$, i.e. $\text{var}(Y \mid -X)$. This means that the joint variation in X and $-X$ contributes to the variation taken in Y at given values of $-X$. As seen above, obtaining the total effect of X by then averaging over the variation in $-X$ summarizes the effect on Y of the interactions involving X, as well as including the first-order effect of X.

In addition to providing a useful interpretation of first-order effects, the Sobol' decomposition equations show that a formal interpretation for an interaction between inputs $X^{(1)}$ and $X^{(2)}$ is

Interpretation 1.3. the additional reduction in the variation in Y expected by fixing both $X^{(1)}$ and $X^{(2)}$ constant (simultaneously), relative to the sum of their individual first-order effects.

This can more usefully be interpreted as the additional variation in Y due to the joint variation in $X^{(1)}$ and $X^{(2)}$, unaccounted for by their individual first-order effects. Similar interpretations can be used for interactions involving more than two inputs.

Given that the total effect of an input is its first-order effect plus any interactions involving the input, the total effect of X can usefully be interpreted as the *total* variation in Y that can be accounted for by the variation in X. This can be seen because an interaction depends on variation in all of the inputs involved, and so averaging over one of these inputs eliminates the interaction. Mathematically, this follows because if an input $X^{(1)}$ does not vary, then

$$E(Y \mid X^{(1)}, X^{(2)}) = E(Y \mid X^{(2)})$$
$$E(Y \mid X^{(1)}) = E(Y)\ ,$$

and similarly for conditional expectations involving more inputs. Plugging these terms into Equation (1.5) to compute the interaction between $X^{(1)}$ and $X^{(2)}$, when

$X^{(1)}$ does not vary, gives

$$\begin{aligned}
&\mathrm{var}_{X^{(1)},X^{(2)}}[E_Y(Y \mid X^{(1)}, X^{(2)})] - \mathrm{var}_{X^{(1)}}[E_Y(Y \mid X^{(1)})] - \mathrm{var}_{X^{(2)}}[E_Y(Y \mid X^{(2)})] \\
&\quad = \mathrm{var}_{X^{(2)}}[E_Y(Y \mid X^{(2)})] - \mathrm{var}[E(Y)] - \mathrm{var}_{X^{(2)}}[E_Y(Y \mid X^{(2)})] \\
&\quad = -\mathrm{var}[E(Y)] \\
&\quad = 0 \, .
\end{aligned}$$

An implication of the total effect of an input $X^{(1)}$ accounting for interactions involving $X^{(1)}$ is that if $X^{(1)}$ only interacts with another input ($X^{(2)}$) then an alternative way to Equation (1.5) of computing the variation in Y due to the interaction between $X^{(1)}$ and $X^{(2)}$ is the total effect of $X^{(1)}$ minus its first-order effect,

$$E_{-X^{(1)}}[\mathrm{var}_Y(Y \mid -X^{(1)})] - \mathrm{var}_{X^{(1)}}[E_Y(Y \mid X^{(1)})] \, .$$

However if $X^{(1)}$ were to be involved in multiple interactions in a mathematical model, then taking the difference between its total effect and its first-order effect would recover all of these interactions that involve $X^{(1)}$. An example of such a mathematical model is $Y = X^{(1)}X^{(2)}X^{(3)}$. Indeed, in this model there is a three-way interaction (between $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$), three two-way interactions (pairwise between $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$), and the first-order effect of each of the three inputs. Each interaction can be isolated using first-order effects, but to isolate the three-way interaction,

$$\begin{aligned}
&\mathrm{var}_{X^{(1)},X^{(2)},X^{(3)}}[E_Y(Y \mid X^{(1)}, X^{(2)}, X^{(3)})] - \mathrm{var}_{X^{(1)},X^{(2)}}[E_Y(Y \mid X^{(1)}, X^{(2)})] \\
&\qquad - \mathrm{var}_{X^{(1)},X^{(3)}}[E_Y(Y \mid X^{(1)}, X^{(3)})] - \mathrm{var}_{X^{(2)},X^{(3)}}[E_Y(Y \mid X^{(2)}, X^{(3)})] \\
&\qquad + \mathrm{var}_{X^{(1)}}[E_Y(Y \mid X^{(1)})] + \mathrm{var}_{X^{(2)}}[E_Y(Y \mid X^{(2)})] + \mathrm{var}_{X^{(3)}}[E_Y(Y \mid X^{(3)})]
\end{aligned}$$

would need to be evaluated. The high number of terms involved in this evaluation is why practioners often focus on only estimating the first-order effect and the total effect of each input (e.g. see Saltelli et al., 2010).

The ideas of this subsection can be summarized as follows. The influence on output Y of each of its inputs can be quantified by first-order effects and total effects of the inputs, called importance measures. The first-order effect ($\mathrm{var}_X[E_Y(Y \mid X)]$) of input X is the reduction in variation in Y expected by fixing X constant. This can more usefully be interpreted as the variation in Y solely due to the variation in X. The total effect ($E_{-X}[\mathrm{var}_Y(Y \mid -X)]$) of X is the reduction in the variation in Y from averaging over X. This can more usefully be interpreted as the total variation in Y that can be accounted for by the variation in X. The total effect of X is its first-order effect plus any interactions between X and other inputs. The

interaction between inputs $X^{(1)}$ and $X^{(2)}$ is the additional reduction in variation in Y expected by fixing both $X^{(1)}$ and $X^{(2)}$ constant (simultaneously), relative to the sum of their individual first-order effects. This can more usefully be interpreted as the additional variation in Y due to the joint variation in $X^{(1)}$ and $X^{(2)}$, unaccounted for by their individual first-order effects. Similar interpretations can be used for interactions involving more than two inputs. If the inputs are independent, then the total variation in output Y can be orthogonally decomposed into the first-order effect of each input and any interactions between the inputs.

1.10.2 Importance measures with correlated inputs

Up to this point the impact on importance measures of dependent inputs has not been discussed. To emphasize the difference between two inputs $X^{(1)}$ and $X^{(2)}$ being dependent from them interacting: interaction between $X^{(1)}$ and $X^{(2)}$ occurs when there is additional variation in Y due to the joint variation in $X^{(1)}$ and $X^{(2)}$, that is unaccounted for by the sum of their individual first-order effects, because of some multiplicative form in the mathematical model (as outlined previously). In contrast, if $X^{(1)}$ and $X^{(2)}$ are dependent then the variation in $X^{(1)}$ affects the variation in $X^{(2)}$, and vice versa, due to some underlying mechanism. An example where $X^{(1)}$ and $X^{(2)}$ are dependent is that there is a physical reason that when $X^{(1)}$ is high, $X^{(2)}$ is also likely to be high. Dependence affects the values taken for $X^{(1)}$ and $X^{(2)}$; interaction affects how these values influence Y.

With independent inputs, it follows from Equation (1.5) that $\text{var}_{X^{(1)},X^{(2)}}[E_Y(Y \mid X^{(1)}, X^{(2)})]$ is equal to the sum of $\text{var}_{X^{(1)}}[E_Y(Y \mid X^{(1)})]$ and $\text{var}_{X^{(2)}}[E_Y(Y \mid X^{(2)})]$ when $X^{(1)}$ and $X^{(2)}$ do not interact. This will not be the case, however, when $X^{(1)}$ and $X^{(2)}$ are dependent (see Saltelli and Tarantola, 2002). To see this, suppose that $X^{(1)}$ and $X^{(2)}$ are the only inputs to a mathematical model, that each have a positive effect on output Y, and that $X^{(1)}$ and $X^{(2)}$ are strongly positively correlated. Now, conditioning on $X^{(1)}$ would be similar to conditioning on both $X^{(1)}$ and $X^{(2)}$ simultaneously, because selecting a value of $X^{(1)}$ means that much is also known about the value of $X^{(2)}$ due to the correlation, and vice versa. As a result, the first-order effect ($\text{var}_{X^{(1)}}[E_Y(Y \mid X^{(1)})]$) of $X^{(1)}$ will be similar to the first-order effect ($\text{var}_{X^{(1)},X^{(2)}}[E_Y(Y \mid X^{(1)}, X^{(2)})]$) of both $X^{(1)}$ and $X^{(2)}$, as will the first-order effect ($\text{var}_{X^{(2)}}[E_Y(Y \mid X^{(2)})]$) of $X^{(2)}$. In this case the two individual first-order effects will sum to greater than the first-order effect of both $X^{(1)}$ and $X^{(2)}$. In other words, each individual first-order effect would be inflated due to the indirect effect of the other input, relative to their sizes in an independent case. Moreover, this would lead to the total effect of $X^{(1)}$ being deflated relative to the independent case, because it is orthogonal to the first-order

effect of $X^{(2)}$ in this two-input case, by the law of total variance (Equation (1.1)), i.e.

$$\text{var}(Y) = \text{var}_{X^{(2)}}[E_Y(Y \mid X^{(2)})] + E_{-X^{(1)}}[\text{var}_Y(Y \mid -X^{(1)})] \, .$$

Correspondingly, the total effect of $X^{(2)}$ would also be deflated relative to the independent case. Indeed, in this case the first-order effect of each input will be greater than its total effect.

Different relationships between the inputs and the output Y than in the simple example above, and different correlation between the inputs, will lead to different effects on the importance measures of dependence between inputs. For example negatively correlated inputs could compensate for each other, deflating the first-order effects and inflating the total effects. Dependence between inputs $X^{(1)}$ and $X^{(2)}$ in a mathematical model will affect the interpretations of their individual importance measures because their individual effects would not be fully disentangled. For example, as well as the variation in $X^{(1)}$, the first-order effect of $X^{(1)}$ could include variation in Y due to the variation in $X^{(2)}$ as a result of the dependence. However the interpretations of the importance measures of any other input (say, $X^{(3)}$) will be unaffected by dependence between $X^{(1)}$ and $X^{(2)}$, and the interpretations of the combined importance measures (e.g. $\text{var}_{X^{(1)},X^{(2)}}[E_Y(Y \mid X^{(1)}, X^{(2)})]$) of $X^{(1)}$ and $X^{(2)}$ will also be unaffected by the dependence. This is because the dependent inputs ($X^{(1)}$ and $X^{(2)}$) are grouped together when conditioning on inputs to compute these importance measures (see Jacques et al., 2006), and so cannot include variation due to that in other inputs such as $X^{(3)}$.

It is because the interpretations of the importance measures for which dependent inputs are grouped together are unaffected by the dependence that Jacques et al. (2006) advocate partitioning the inputs into groups that can have within-group, but not between-group, correlation, and using the law of total variance to obtain the importance measures for each group. Where such groups are appropriate, this strategy will circumvent the effect of dependence between inputs on the interpretations of importance measures. However, it also restricts the analysis to only quantifying the influence on Y of groups of inputs, providing no information on the relative importance of each input within a group, which lead to Xu and Gertner (2007) and Da Veiga et al. (2009) suggesting alternative methods in the presence of correlated inputs.

Despite the complication of dependent inputs, Saltelli and Tarantola (2002) and Saltelli et al. (2010) highlight that interpreting an importance measure as a reduction in the variation in the output holds in the presence of correlated inputs. For example the first-order effect of input X is formally defined as the reduction in variation

in output Y expected by fixing X constant (Interpretation 1.1), and this is the case whether the reduction is due to the direct effect of fixing X or due to the indirect effect from another input as a result of dependence shared with X. This is because importance measures are still defined by $\text{var.}[E_Y(Y \mid \cdot)]$ or $E.[\text{var}_Y(Y \mid \cdot)]$, and so have Interpretations 1.1 and 1.2, in the presence of correlated inputs. The complication of dependent inputs is that whereas we can infer that a reduction in variation in the output is due to the direct effect of the input being fixed in the independent case, this will not necessarily be the case in the presence of correlated inputs. However Interpretations 1.1 and 1.2 are meaningful in themselves, and so insight into the importance of each input can still be gained from the individual importance measures of correlated inputs. This is particularly the case when ranking the importance of inputs, for example if the importance measures of an input X are substantially greater than those of other inputs (which are all dependent), then this would indicate that the variation in X is more important to the output than that of the other inputs. We will discuss the impact of correlated inputs in the context of analysing the representation of a physical process by climate models in Chapter 2.

1.10.3 Computing importance measures

Importance measures can be computed through random number simulation (see Jansen, 1999). Assuming that PDFs can be specified for the inputs, which may include specifying a dependence structure between them, then simulation can be used to compute any importance measure. As an example, the following outlines the simulation procedure for obtaining the first-order effect ($\text{var}_X[E_Y(Y \mid X)]$) of input X on output Y of a mathematical model, where other inputs are also present in the model:

- Randomly generate a value for X from its specified PDF.
- Randomly generate many values for each of the other inputs from their specified joint PDF, given the value for X, to explore the (conditional) joint PDF of these inputs.
- Use these values with the specification of the mathematical model to compute a corresponding set of values for Y.
- Compute the mean of the set of values for Y, $E_Y(Y \mid X)$.
- Repeat the previous steps many times to explore the PDF of X, obtaining a set of values for $E_Y(Y \mid X)$ for the different values for X.
- Compute the variance of the set of values for $E_Y(Y \mid X)$, $\text{var}_X[E_Y(Y \mid X)]$.

To simulate the random numbers required to compute each term, ordinary random sampling is described as "the obvious choice" by Jansen (1999), although alternatives include Latin hypercube sampling (see McKay et al., 2000) or quasi-random sampling (see Sobol', 1976). These can be more efficient than ordinary random sampling in different scenarios (Jansen, 1999; Homma and Saltelli, 1996; Saltelli et al., 2010).

In the literature referenced in this section, the importance measures are generally normalized by the total variation in output Y (e.g. see Saltelli et al., 2010). To estimate the total variation in Y, the model specified for the inputs is used with the mathematical model to simulate samples of Y. Computing the variance of these samples provides the estimate for the total variation in Y.

1.10.4 Uses for importance measures

Normalized importance measures have been used and demonstrated with applications such as Saltelli and Tarantola (2002) analysing the E-level mathematical model, which is used to compare the performance of computer codes used in safety assessment for nuclear waste disposal (also see Homma and Saltelli, 1996). Saltelli and Tarantola determined the input that accounts for the most variation in the output, and the smallest set of inputs that can be fixed constant to reduce the variation in the output by a predetermined amount. Expert judgement and physical theory were used to specify the distributions of the inputs in the E-level model.

As well as established mathematical models such as the E-level model, importance measures have been used to study climate change. Hof et al. (2008) analysed the impact of scientific uncertainties in future climate and value judgements in climate policy. Anderson et al. (2014) applied a global sensitivity analysis to the DICE model of Nordhaus (2008) for global warming. The DICE model is an integrated assessment model that examines the interplay between economics, energy use and climate change. Amongst other things, the authors make the case for using global sensitivity analyses to identify the greatest sources of uncertainty in the modelling of climate change.

The focus of the literature referenced in this section is on deterministic mathematical models (i.e. mathematical models without a residual term). Despite this, the application of the global sensitivity analysis to statistical models is immediate: the response variable of the statistical model is its output, and the inputs are the predictor variables and random effect terms, in addition to the residual term. Indeed,

a benefit of performing a global sensitivity analysis with a statistical model is that the parameters of the model are estimated using empirical data, and so the data can be used when specifying the model for the inputs. Uncertainty in the parameters of a statistical model is represented by standard errors (in a frequentist setting) or posterior distributions (in a Bayesian setting). The impact of this uncertainty could be investigated with a global sensitivity analysis by including the parameters of the statistical model as inputs. Gelman and Pardoe (2007) discuss this in the context of their predictive comparison measures.

Our focus is on GLMMs, and in Section 1.9 we discussed the sources of variation in a GLMM used to capture the representation of a driving physical process of precipitation in climate models. Recall that these sources of variation are 1) the temporal variation in each predictor, 2) the variation across climate models in the distribution of each predictor, 3) the variation across climate models in the relationship between each predictor and the precipitation, 4) further variation across climate models in precipitation bias, and 5) the further internal variation in the precipitation that is unaccounted for by the physical process or precipitation bias. Focusing on these sources of variation in a GLMM (rather than the uncertainty in the parameters of the GLMM), importance measures can be used to quantify the extent to which, and understand why, the representation of the physical process varies across and within the climate models.

In Chapter 2 we propose a general methodology for performing a global sensitivity analysis with GLMMs in the context of quantifying the extent to which the representation of driving physical processes of precipitation vary across and within climate models. A general interpretation is given for different importance measures in this context, as well as a recommended model for the inputs. Additionally, the use of importance measures is demonstrated with an analytical example and a numerical example. In Chapters 3 to 6 we use the methodology to analyse the simulation of precipitation in two different ensembles of climate models, and on various different spatial and temporal scales. In Chapter 6 this includes extending the methodology of Chapter 2 to focus on extreme precipitation. We then provide a summary and final discussion for the thesis in Chapter 7.

2 Methodology

Aim: To propose a methodology to quantify the extent to which, and understand why, the representation of driving physical processes of precipitation varies across and within climate models.

2.1 Global sensitivity analysis with GLMMs: Analytical example

Suppose that a generalized linear mixed model (GLMM) captures the representation of a driving physical process of precipitation in an ensemble of climate models on a given spatial and temporal scale. In our GLMM, predictor variables are used to capture the representation of the atmospheric conditions of interest by the climate models, and these represented atmospheric conditions and their estimated relationship with precipitation capture the representation of the physical process of interest by the climate models. Ensemble member is treated as a random factor, and as such random effects are used to allow the estimated relationship between the represented atmospheric conditions and precipitation to vary with the ensemble member. This is done by using the random effects to allow the relationship between precipitation and each predictor to vary across the ensemble members, as well as allowing variation across the members in any additional bias in precipitation. Specifically, we define variation in precipitation bias as variation across climate models in their simulated mean rate of precipitation.

As seen in Subsection 1.10.1, when the inputs to a mathematical model are independent the law of total variance can be used to decompose the total variation in the output of the model into the first-order effect of each input and interactions between them. This is called a global sensitivity analysis. As we outline in this chapter, a global sensitivity analysis can be used with our GLMM to decompose the variation in precipitation into the effects of the temporal variation in the predictors, of the variation in the distributions of the predictors across ensemble members, and of the variation in the relationships between the predictors and precipitation across ensemble members. In doing so, we can quantify the extent to which, and understand why, the representation of driving physical processes of precipitation vary across and within climate models. Before outlining the decomposition in the general case, we use an analytical example to demonstrate the concepts.

To demonstrate the use of importance measures with statistical models, we analytically decompose the variation in a response variable using a simple statistical

model specification. We also decompose importance measures and compare them to this decomposition of the variation in the response variable. Let i index time and j index the member of a climate model ensemble, let X_{ij} be a predictor, $\psi_j^{(0)}$ and $\psi_j^{(1)}$ be random effects, and ϵ_{ij} be the residual term. The residual term is the difference between the response Y_{ij} and its fitted expectation μ_{ij}. The random effects allow variation across the members of the ensemble, whereas predictor X_{ij} varies with both time and ensemble member. Suppose that these inputs relate to Y_{ij} by

$$\begin{aligned} Y_{ij} &= \mu_{ij} + \epsilon_{ij} \\ &= \beta_0 + \beta_1 X_{ij} + \psi_j^{(0)} + \psi_j^{(1)} X_{ij} + \epsilon_{ij} \ , \end{aligned} \tag{2.1}$$

where β_0 and β_1 are fixed parameters.

Before decomposing the variation in response Y_{ij}, it is worth using the specification above to consider the role of the intercept of GLMMs. The intercept of a GLMM is defined to be the fitted value (μ_{ij}) when each of the predictors equal zero, i.e. $X_{ij} = 0$ in this specification. Given this, the random effect $\psi_j^{(0)}$ captures the variation across ensemble members in the intercept. As such, the first-order effect of $\psi_j^{(0)}$ is only of interest if the intercept of the statistical model is physically meaningful and of interest, which will not always be the case. To see this suppose that X_{ij} represents temperature. In this scenario the physical interpretation of the intercept would depend on whether X_{ij} is measured in Kelvin or degrees Celsius. Measuring X_{ij} in degrees Celsius would mean that the scale of X_{ij} is somewhat arbitrary, such that $X_{ij} = 0$ does not mean the absence of heat, whereas measuring X_{ij} in Kelvin may mean that $X_{ij} = 0$ is so far from its empirical distribution that it is not relevant to the analysis. Moreover predictors are often transformed onto different scales before being used in a GLMM to (e.g.) capture non-linear relationships between the predictors and the response. This could change the physical interpretation of the intercept.

As shown later, in addition to the first-order effect of $\psi_j^{(0)}$, the scale on which the predictors are used in a GLMM affects the interpretation of the first-order effect of $\psi_j^{(1)}$. The result is that, in addition to the first-order effect of $\psi_j^{(0)}$, the first-order effect of $\psi_j^{(1)}$ may not be of interest. However it can be seen that the combined first-order effect of the random effects is physically meaningful in the general case. This follows from its mathematical definition, $\text{var}_{\boldsymbol{\psi}}[E_Y(Y \mid \boldsymbol{\psi})]$ (where $\boldsymbol{\psi}$ denotes all random effects in a GLMM), since $\text{var}_{\boldsymbol{\psi}}[E_Y(Y \mid \boldsymbol{\psi})]$ quantifies the variation in precipitation bias (i.e. mean precipitation rate) remaining after averaging over the variation in the predictors. As a result, we advocate that the first-order effect of all of the random effects together is generally estimated, but that it is not decomposed into the individual contribution of each random effect unless their individual effects

are physically meaningful and of interest in the application. As we show in the next section, the interpretations of the other importance measures are physically meaningful in the general case.

Returning to decomposing the variation in Y_{ij} in Equation (2.1), we are interested in separating temporal variation from variation that is due to the differences between ensemble members. To this end it is worth decomposing the variation in predictor X_{ij}. This is because the differences between ensemble members could cause variation across the members in the distribution of a predictor. For example there could be bias in the mean of a predictor in a climate model, and this could affect its simulation of precipitation. To this end, the variation in X_{ij} can be decomposed into the variation across ensemble members in its distribution, and its remaining (temporal) variation. A first-order approximation to this is to remove the mean from X_{ij} within each ensemble member, and consider these within-member means as a separate input to the statistical model, along with the temporal variation in X_{ij} within each member. Let

$$X_{ij} = \gamma_j + X'_{ij} \ ,$$

where γ_j denotes the mean within the jth ensemble member of X_{ij}, and X'_{ij} denotes its deviation from these means. Therefore X'_{ij} is centred on zero within each ensemble member, and γ_j allows for different means of the predictor in different ensemble members.

The inputs are now X'_{ij}, γ_j, $\psi_j^{(1)}$, $\psi_j^{(0)}$, as well as the residual term ϵ_{ij}. Let these inputs be independent in this example, and let their means and variances be

$$\begin{aligned} \psi_j^{(0)} &\sim (0, \tau_0) & \gamma_j &\sim (\mu_\gamma, \tau_\gamma) \\ \psi_j^{(1)} &\sim (0, \tau_1) & X'_{ij} &\sim (0, \tau_{X'}) \\ & & \epsilon_{ij} &\sim (0, \tau_\epsilon) \ . \end{aligned}$$

To clarify terminology used below and throughout the thesis, we label the inputs that capture variation across ensemble members in the intercept or relationships between predictors and the response (i.e. $\psi_j^{(0)}$ and $\psi_j^{(1)}$ in this example) as *random effects* (or random effect terms). This excludes the inputs (X'_{ij} and γ_j) that capture variation in the predictor variables, and also excludes the residual term (ϵ_{ij}). This terminology is consistent with that used in Section 1.9 when defining the GLMM class of statistical models.

Excluding subscripts for brevity, the variation in Y can be analytically decomposed in this example because there are no features such as dependence between the inputs

or a non-linear link function that could complicate the decomposition. The total variation in Y is

$$\begin{aligned}
&\text{var}(Y) \\
&\quad = \text{var}[\beta_0 + \beta_1(\gamma + X') + \psi^{(0)} + \psi^{(1)}(\gamma + X') + \epsilon] \\
&\quad = \beta_1^2\text{var}(\gamma) + \beta_1^2\text{var}(X') + \tau_0 + \text{var}(\psi^{(1)}\gamma) + \text{var}(\psi^{(1)}X') + \tau_\epsilon \\
&\quad = \beta_1^2\tau_\gamma + \beta_1^2\tau_{X'} + \tau_0 + E[(\psi^{(1)})^2]E(\gamma^2) - E(\psi^{(1)})^2E(\gamma)^2 \\
&\qquad + E[(\psi^{(1)})^2]E(X'^2) - E(\psi^{(1)})^2E(X')^2 + \tau_\epsilon \\
&\quad = \beta_1^2\tau_\gamma + \beta_1^2\tau_{X'} + \tau_0 + \text{var}(\psi^{(1)})[\text{var}(\gamma) + E(\gamma)^2] \\
&\qquad + \text{var}(\psi^{(1)})\text{var}(X') + \tau_\epsilon \\
&\quad = \tau_0 + \tau_1\mu_\gamma^2 + \beta_1^2\tau_\gamma + \beta_1^2\tau_{X'} + \tau_1\tau_\gamma + \tau_1\tau_{X'} + \tau_\epsilon \ . \qquad (2.2)
\end{aligned}$$

In this statistical model specification there are two (bivariate) multiplicative terms between different inputs, that between X' and $\psi^{(1)}$ and that between γ and $\psi^{(1)}$, and so interactions between these inputs are the only two interactions present. Therefore as we show below, Equation (2.2) decomposes the total variation in Y into the first-order effect (FOE) of each input and the two interactions (int),

$$\begin{aligned}
\text{var}(Y) = \text{FOE}(\psi^{(0)}) + \text{FOE}(\psi^{(1)}) + \text{FOE}(\gamma) + \text{FOE}(X') + \text{int}(\gamma, \psi^{(1)}) \\
+ \text{int}(X', \psi^{(1)}) + \text{FOE}(\epsilon) \ . \qquad (2.3)
\end{aligned}$$

As such, each of the terms in Equation (2.2) can be recovered using importance measures. To see how the importance measures can be analytically decomposed in this example, consider the first-order effect of $\psi^{(1)}$,

$$\begin{aligned}
&\text{var}_{\psi^{(1)}}[E_Y(Y \mid \psi^{(1)})] \\
&\quad = \text{var}\{E(\beta_0 \mid \psi^{(1)}) + E[\beta_1(\gamma + X') \mid \psi^{(1)}] + E(\psi^{(0)} \mid \psi^{(1)}) \\
&\qquad + E[\psi^{(1)}(\gamma + X') \mid \psi^{(1)}] + E(\epsilon \mid \psi^{(1)})\} \\
&\quad = \text{var}[\beta_0 + \beta_1 E(\gamma + X') + E(\psi^{(0)}) + \psi^{(1)}E(\gamma + X')] \\
&\quad = \text{var}[\psi^{(1)}E(\gamma + X')] \\
&\quad = E(\gamma)^2\text{var}(\psi^{(1)}) \\
&\quad = \tau_1\mu_\gamma^2 \ .
\end{aligned}$$

This is the second term in Equation (2.2). The input $\psi^{(1)}$ captures the variation in the relationship between X and Y across ensemble members, and so its first-

order effect is the variation in Y solely due to this variation in the relationship between X and Y. As discussed above, however, this term depends on the scale on which predictor X is used in the statistical model, and this can be seen here because $\text{FOE}(\psi^{(1)})$ depends on the overall mean (μ_γ) of X. As such, $\text{FOE}(\psi^{(1)})$ may only be physically meaningful and of interest when considered as a part of the combined first-order effect of both random effects ($\text{FOE}(\psi^{(0)}, \psi^{(1)})$). This combined first-order effect is the variation in precipitation bias remaining after averaging over the variation in X.

In addition to the second term in Equation (2.2), the other sources of variation can be recovered and interpreted using importance measures:

- $\text{FOE}(\psi^{(0)}) = \tau_0$, the variation solely due to the variation across ensemble members in the intercept.
- $\text{FOE}(\psi^{(1)}) = \tau_1 \mu_\gamma^2$, the variation solely due to the variation across ensemble members in the relationship between X and Y.
- $\text{FOE}(\gamma) = \beta_1^2 \tau_\gamma$, the variation solely due to the variation across ensemble members in the distribution (mean) of X.
- $\text{FOE}(X') = \beta_1^2 \tau_{X'}$, the variation solely due to the temporal variation in X.
- $\text{int}(\gamma, \psi^{(1)}) = \tau_1 \tau_\gamma$, the additional variation due to the joint variation in the distribution (mean) of X and its relationship with Y, unaccounted for by their individual first-order effects.
- $\text{int}(X', \psi^{(1)}) = \tau_1 \tau_{X'}$, the additional variation in Y due to the joint variation in the temporal component of X and its relationship with Y, unaccounted for by their individual first-order effects.
- $\text{FOE}(\epsilon) = \tau_\epsilon$, the residual (unaccounted for) variation in Y.

As outlined in Subsection 1.10.1, each interaction ($\text{int}(\gamma, \psi^{(1)})$ and $\text{int}(X', \psi^{(1)})$) can be computed by the difference between the appropriate first-order effects, e.g.

$$\text{int}(\gamma, \psi^{(1)}) = \text{FOE}(\gamma, \psi^{(1)}) - \text{FOE}(\gamma) - \text{FOE}(\psi^{(1)}) \tag{2.4}$$

would isolate $\text{int}(\gamma, \psi^{(1)}) = \tau_1 \tau_\gamma$.

In addition to Equation (2.4), $\text{int}(\gamma, \psi^{(1)})$ can be computed by using the total effect ($\text{TE}(\gamma)$) of γ,

$$\text{int}(\gamma, \psi^{(1)}) = \text{TE}(\gamma) - \text{FOE}(\gamma) \; . \tag{2.5}$$

Equation (2.5) holds in this example because γ only interacts with $\psi^{(1)}$. Computing the difference between TE($\psi^{(1)}$) and FOE($\psi^{(1)}$) would isolate both interactions present in the statistical model. To show this, consider decomposing the total effect of $\psi^{(1)}$ (using the notation outlined in Subsection 1.10.1 to denote all inputs (including residual term ϵ) except for $\psi^{(1)}$ as $-\psi^{(1)}$),

$$
\begin{aligned}
&E_{-\psi^{(1)}}[\mathrm{var}_Y(Y \mid -\psi^{(1)})] \\
&\quad = E\{\mathrm{var}[\beta_0 + \beta_1(\gamma + X') + \psi^{(0)} + \psi^{(1)}(\gamma + X') + \epsilon \mid -\psi^{(1)}]\} \\
&\quad = E\{\mathrm{var}[\psi^{(1)}(\gamma + X') \mid -\psi^{(1)}]\} \\
&\quad = E[(\gamma + X')^2 \mathrm{var}(\psi^{(1)})] \\
&\quad = E[(\gamma + X')^2]\tau_1 \\
&\quad = [\mathrm{var}(\gamma + X') + E(\gamma + X')^2]\tau_1 \\
&\quad = [\mathrm{var}(\gamma) + \mathrm{var}(X') + E(\gamma)^2]\tau_1 \\
&\quad = \tau_1\mu_\gamma^2 + \tau_1\tau_\gamma + \tau_1\tau_{X'} \, .
\end{aligned}
$$

This includes the variation in Y due to the interactions ($\tau_1\tau_\gamma$ and $\tau_1\tau_{X'}$) between X and its relationship with Y, i.e. both interactions shown in Equation (2.2). Computing the difference between the total effect of $\psi^{(1)}$ and its first-order effect provides these interactions. Note that this is in contrast to $\psi^{(0)}$ and ϵ because neither of these inputs are involved in any interactions and so their first-order effects are equal to their total effects.

Given the discussion in this section, we can summarize the sources of variation in response Y captured by the statistical model of Equation (2.1) as follows. The expectation of Y depends on X and on the ensemble member, and this is captured by a statistical model. Predictor X varies with time and also its distribution varies with ensemble member. The variation in Y solely due to the temporal variation in X is quantified by the first-order effect of X', and the variation solely due to the variation in the distribution of X across the ensemble members is quantified by the first-order effect of γ. If the variation in X is averaged over, the remaining variation in precipitation bias is quantified by the first-order effect of $\psi^{(0)}$ and $\psi^{(1)}$. The variation due to the differences between ensemble members can also depend on the value taken for X. This is because X interacts with its relationship with Y, and the variation due to these interactions is quantified by the interaction between X' and $\psi^{(1)}$ and by that between γ and $\psi^{(1)}$. However the variables X, $\psi^{(0)}$ and $\psi^{(1)}$ are inputs to a statistical model and so will not account for all of the variation in Y. The remaining variation in Y is the residual variation, which can be quantified by either the first-order effect of ϵ or by its total effect.

Before leaving this analytical example to outline a similar decomposition in the general case, it is worth examining the sources of variation captured by the first-order effect (FOE($\boldsymbol{\psi}$)) of the random effects in more detail. Recall that this is the variation in precipitation bias remaining after averaging over the variation in the predictors, but that it will not always be physically meaningful and useful to decompose FOE($\boldsymbol{\psi}$) into the individual contribution of each random effect. In the example used in this section,

$$\text{FOE}(\boldsymbol{\psi}) = \text{FOE}(\psi^{(0)}, \psi^{(1)}) = \tau_0 + \tau_1 \mu_\gamma^2 \, ,$$

where τ_0 and τ_1 are the variance components of $\psi^{(0)}$ and $\psi^{(1)}$ respectively, and μ_γ is the overall mean of predictor X. As such, FOE($\boldsymbol{\psi}$) includes variation due to that across ensemble members in the relationship between X and Y, and the additional variation in the mean rate of precipitation across the members.

As stated previously, we define variation in mean precipitation rate to be variation across ensemble members in precipitation bias. In this example X is used to capture the representation of the atmospheric conditions of interest, and these represented atmospheric conditions and their estimated relationship with precipitation capture the representation of the physical process of interest. We are therefore interested in the precipitation bias that results from the variation across members in the representation of this physical process. However precipitation bias can also be present because of variation across the ensemble members in their representation of other physical processes, which will not necessarily be of interest. This is because the differences between ensemble members will generally cause variation across the members not only in the representation of the physical process of interest, but also in the representation of other physical processes relevant to precipitation, and this will generally lead to some variation across ensemble members in the mean rate of precipitation, i.e. precipitation bias.

It is a limitation of the methodology that the precipitation bias due to the physical process of interest often cannot be fully disentangled from the precipitation bias that is due to other physical processes. Often all that can be said is that this precipitation bias due to other physical processes is included in the first-order effect (FOE($\boldsymbol{\psi}$)) of the random effects. This precipitation bias due to other physical processes could be defined as the variation across ensemble members in mean precipitation rate in the absence of the atmospheric conditions captured by X by using the random effect ($\psi^{(0)}$) associated with the intercept. However as discussed previously, the intercept is often either somewhat arbitrary or unrealistic. We used temperature as an example previously, but another example is when X measures the amount

of moisture in the atmosphere, perhaps the global-mean annual-mean humidity. In this case X could never realistically be zero. This further emphasizes why we can define the first-order effect of the random effects as the variation across ensemble members in precipitation bias remaining after averaging over the variation in X, but it will often not be useful to decompose this into the individual contribution of each random effect.

2.2 Global sensitivity analysis with GLMMs

2.2.1 General decomposition structure

In the previous section we used an analytical example to demonstrate the concepts behind applying a global sensitivity analysis to a statistical model capturing the representation of a driving physical process of precipitation by an ensemble of climate models. In such a statistical model, random effects are used to allow variation across ensemble members in the relationships between predictors and the precipitation. In this section we outline the decomposition structure in the general case for such a GLMM. In doing so we provide a general structure for the decomposition, and an interpretation for the different importance measures. We assume independence between all of the predictors and random effects in this subsection (although we do discuss dependence between the residual term of a GLMM and the other inputs), before discussing the impact of dependence between predictors and random effects in the succeeding sections. We present the general decomposition as hierarchical. We do this because the full decomposition may not be required for any given application, and because the dependence between inputs that we discuss in succeeding sections may only affect an application at lower levels of the hierarchical structure.

Let Y be precipitation, μ be its fitted expectation from our GLMM, and ϵ be the residual term, such that

$$Y = \mu + \epsilon \, ,$$

where the expectation of ϵ is zero. Let μ depend on a set of predictors ($\boldsymbol{X}$) and a set of random effects ($\boldsymbol{\psi}$), as well as fixed parameters that are estimated when the GLMM is fitted to data. Following the decomposition of each predictor outlined in Section 2.1, we decompose the predictors into their means ($\boldsymbol{\gamma}$) within each ensemble member and their temporal variation ($\boldsymbol{X'}$) from these means, $\boldsymbol{X} = \boldsymbol{\gamma} + \boldsymbol{X'}$. Therefore,

$$\mu = \mu(\boldsymbol{X'}, \boldsymbol{\gamma}, \boldsymbol{\psi}) \, ,$$

is a function of the inputs $\boldsymbol{X'}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$. We assume that

$$E(\epsilon \mid \boldsymbol{X'}, \boldsymbol{\gamma}, \boldsymbol{\psi}) = 0 \tag{2.6}$$

for all $\boldsymbol{X'}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$. This implies that

$$\begin{aligned} E(\epsilon \mid \mu) &= E[E(\epsilon \mid \mu, \boldsymbol{X'}, \boldsymbol{\gamma}, \boldsymbol{\psi}) \mid \mu] \\ &= E[E(\epsilon \mid \boldsymbol{X'}, \boldsymbol{\gamma}, \boldsymbol{\psi}) \mid \mu] \\ &= 0 \ , \end{aligned} \tag{2.7}$$

and so

$$E_Y(Y \mid \mu) = E_Y(\mu + \epsilon \mid \mu) = \mu \ . \tag{2.8}$$

Recalling Subsection 1.10.1, the law of total variance can be used to decompose the total variation in precipitation by quantifying the variation solely due to the variation in an input (its first-order effect) and the remaining variation in the precipitation (the total effect of all other inputs). By initally considering the expectation μ, and the residual term ϵ, as the two inputs to our GLMM, we have

$$\begin{aligned} \text{var}(Y) &= \text{var}_\mu[E_Y(Y \mid \mu)] + E_\mu[\text{var}_Y(Y \mid \mu)] \\ &= \text{var}_\mu[E_Y(Y \mid \mu)] + E_{-\epsilon}[\text{var}_Y(Y \mid -\epsilon)] \\ &= \text{FOE}(\mu) + \text{TE}(\epsilon) \ . \end{aligned} \tag{2.9}$$

Given that $Y = \mu + \epsilon$, the inputs μ and ϵ cannot interact because they do not combine in any multiplicative way, but recall from Subsection 1.10.2 that dependence between inputs can affect the estimation of importance measures, and therefore we investigate below whether or not dependence between μ and ϵ could affect the estimation of the decomposition of Equation (2.9). Such dependence between μ and ϵ could occur when the shape (e.g. variance) of the distribution fitted by the GLMM varies, for example because the fitted dispersion parameter λ of a GLMM is allowed to vary with inputs (e.g. $\boldsymbol{X'}$ or $\boldsymbol{\gamma}$ defined above), or because there is a relationship between the expectation μ of the fitted distribution and the shape of the distribution built into the family of distributions assumed for the precipitation (e.g. when the Gamma family is assumed).

Although μ and ϵ may be dependent, Equation (2.8) implies that

$$\text{FOE}(\mu) = \text{var}_\mu[E_Y(Y \mid \mu)] = \text{var}(\mu) \ ,$$

and Equation (2.7) implies that

$$\begin{aligned}\mathrm{TE}(\epsilon) &= E_\mu[\mathrm{var}_Y(Y \mid \mu)] \\ &= E_\mu[\mathrm{var}(\epsilon \mid \mu)] \\ &= \mathrm{var}(\epsilon) - \mathrm{var}[E(\epsilon \mid \mu)] \\ &= \mathrm{var}(\epsilon)\ .\end{aligned}$$

Therefore, Equation (2.9) can be written as

$$\mathrm{var}(Y) = \mathrm{var}(\mu) + \mathrm{var}(\epsilon)\ . \tag{2.10}$$

Our assumption of Equation (2.6) implies that, although μ and ϵ may be dependent, such that the inputs to μ can affect the shape of the distribution of ϵ, the expectation of ϵ is always zero and so μ and ϵ are uncorrelated.

With respect to interpretation, we can define the first term ($\mathrm{var}(\mu) = \mathrm{FOE}(\mu)$) of Equation (2.10) as the variation in precipitation accounted for by our GLMM, or

Interpretation 2.1. the variation in precipitation accounted for by the variation across and within the ensemble members in the representation of the physical process and precipitation bias.

The second term ($\mathrm{TE}(\epsilon)$) of Equation (2.10) is the remaining residual variation, or

Interpretation 2.2. the variation in precipitation unaccounted for by the representation of the physical process and precipitation bias.

As stated above, we are presenting a hierarchical decomposition in this section. Therefore we can summarize our general decomposition using a diagram, with Equation (2.10) and Interpretations 2.1 and 2.2 as the starting point. This starting point is shown in Figure 2.1, and the diagram will be developed as the decomposition progresses. The end point of the decomposition will be the total decomposition of the variation in precipitation into the first-order effects of the inputs and interactions between them (similar to Equation (2.3), page 46, in Section 2.1). However, some terms at a higher level in the hierarchical decomposition are particularly useful, and so obtaining these will guide our decomposition.

Equation (2.10) can be developed by applying the law of total variance for a second time, this time to decompose the total variation in μ. Since μ is a function of the temporal components ($\boldsymbol{X'}$) of the predictors, their means ($\boldsymbol{\gamma}$) within each ensemble member, and the random effects ($\boldsymbol{\psi}$), the input conditioned on in this second application of the law of total variance could be any set or combination of these

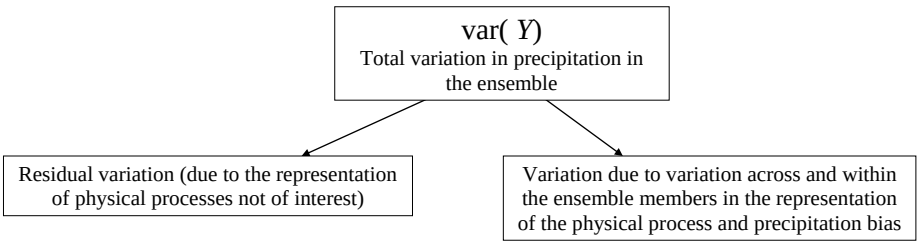

Figure 2.1: First part of the proposed decomposition structure for precipitation (Y) in an ensemble of climate models.

inputs. The predictor means and random effects capture variation across different ensemble members, and so differ with respect to the temporal components of the predictors. Therefore let the variation in μ be decomposed by conditioning on the temporal components ($\boldsymbol{X'}$) of the predictors,

$$\begin{aligned}\text{var}(\mu) &= \text{var}_{\boldsymbol{X'}}[E_\mu(\mu \mid \boldsymbol{X'})] + E_{\boldsymbol{X'}}[\text{var}_\mu(\mu \mid \boldsymbol{X'})] \\ &= \text{var}_{\boldsymbol{X'}}[E_\mu(\mu \mid \boldsymbol{X'})] + E_{-(\boldsymbol{\gamma},\boldsymbol{\psi})}\{\text{var}_\mu[\mu \mid -(\boldsymbol{\gamma},\boldsymbol{\psi})]\} \\ &= \text{FOE}_\mu(\boldsymbol{X'}) + \text{TE}_\mu(\boldsymbol{\gamma},\boldsymbol{\psi}) \ , \end{aligned} \tag{2.11}$$

where FOE_μ and TE_μ denote first-order effect and total effect on expectation μ respectively. This holds because the predictor inputs ($\boldsymbol{X'}$ and $\boldsymbol{\gamma}$) and the random effects ($\boldsymbol{\psi}$) are the only inputs to μ, and so another way of viewing the set of inputs $\boldsymbol{X'}$ is as the set of all inputs to μ except for the inputs $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$, i.e. the set $-(\boldsymbol{\gamma},\boldsymbol{\psi})$.

Equation (2.11) provides a first-order effect and a total effect on the variation in μ. First-order effects on μ are, however, equivalent to first-order effects on Y because

$$E(\epsilon \mid \boldsymbol{X'}) = E[E(\epsilon \mid \boldsymbol{X'},\boldsymbol{\gamma},\boldsymbol{\psi}) \mid \boldsymbol{X'}] = 0 \tag{2.12}$$

by the assumption of Equation (2.6), and so

$$\begin{aligned} E_Y(Y \mid \boldsymbol{X'}) &= E_\mu(\mu \mid \boldsymbol{X'}) + E(\epsilon \mid \boldsymbol{X'}) \\ &= E_\mu(\mu \mid \boldsymbol{X'}) \ , \end{aligned}$$

and hence $\text{var}_{\boldsymbol{X'}}[E_\mu(\mu \mid \boldsymbol{X'})] = \text{var}_{\boldsymbol{X'}}[E_Y(Y \mid \boldsymbol{X'})]$. Moreover, it therefore holds that interactions are also equivalent when decomposing the variation in μ and Y because interactions are defined as the difference between different first-order effects (recall Equation (1.5), page 34).

Although first-order effects on μ are equivalent to those on Y, this is not necessarily the case with total effects. As noted in Subsection 1.10.2, dependence between inputs can affect the estimation of first-order effects and total effects differently,

and, as we highlight below, dependence between the shape of the distribution of the residual term ϵ and other inputs can result in total effects on μ being different to those on Y.

Recall that $\boldsymbol{X'}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\psi}$ are the inputs to μ, and that $\boldsymbol{X'}$, $\boldsymbol{\gamma}$, $\boldsymbol{\psi}$, and ϵ are the inputs to Y. We have

$$\begin{aligned}\mathrm{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi}) &= E[\mathrm{var}(Y \mid \boldsymbol{X'}, \epsilon)] \\ &= E[\mathrm{var}(\mu + \epsilon \mid \boldsymbol{X'}, \epsilon)] \\ &= E[\mathrm{var}(\mu \mid \boldsymbol{X'}, \epsilon)] \ .\end{aligned}$$

This equals $\mathrm{TE}_\mu(\boldsymbol{\gamma}, \boldsymbol{\psi}) = E[\mathrm{var}(\mu \mid \boldsymbol{X'})]$ if μ and ϵ are conditionally independent given $\boldsymbol{X'}$, but typically differs from $\mathrm{TE}_\mu(\boldsymbol{\gamma}, \boldsymbol{\psi})$ if μ and ϵ depend on $\boldsymbol{X'}$ (either directly or via dependence with other inputs). It is true, however, that $\mathrm{TE}_\mu(\boldsymbol{\gamma}, \boldsymbol{\psi}) \geq \mathrm{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$. To see this, note that the law of total variance yields

$$\mathrm{var}(\mu \mid \boldsymbol{X'}) = E[\mathrm{var}(\mu \mid \boldsymbol{X'}, \epsilon) \mid \boldsymbol{X'}] + \mathrm{var}[E(\mu \mid \boldsymbol{X'}, \epsilon) \mid \boldsymbol{X'}] \ ,$$

and so

$$\begin{aligned}\mathrm{TE}_\mu(\boldsymbol{\gamma}, \boldsymbol{\psi}) &= E[\mathrm{var}(\mu \mid \boldsymbol{X'})] \\ &= E\{E[\mathrm{var}(\mu \mid \boldsymbol{X'}, \epsilon) \mid \boldsymbol{X'}]\} + E\{\mathrm{var}[E(\mu \mid \boldsymbol{X'}, \epsilon) \mid \boldsymbol{X'}]\} \\ &= E[\mathrm{var}(\mu \mid \boldsymbol{X'}, \epsilon)] + E\{\mathrm{var}[E(\mu \mid \boldsymbol{X'}, \epsilon) \mid \boldsymbol{X'}]\} \\ &\geq E[\mathrm{var}(\mu \mid \boldsymbol{X'}, \epsilon)] \\ &= \mathrm{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi}) \ .\end{aligned}$$

Recapping, we have shown that

$$\begin{aligned}\mathrm{var}(Y) &= \mathrm{var}(\mu) + \mathrm{var}(\epsilon) \\ &= \mathrm{FOE}(\boldsymbol{X'}) + \mathrm{TE}_\mu(\boldsymbol{\gamma}, \boldsymbol{\psi}) + \mathrm{var}(\epsilon) \ ,\end{aligned}$$

where $\mathrm{TE}_\mu(\boldsymbol{\gamma}, \boldsymbol{\psi})$ quantifies the total contribution of $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$ to the explained variation ($\mathrm{var}(\mu)$), and can be used to summarize the first-order effects of $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$, and the interactions involving them. More than this however, as we discuss below, $\mathrm{TE}_\mu(\boldsymbol{\gamma}, \boldsymbol{\psi})$ can be interpreted as the a more reliable measure of the total effect on precipitation of $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$ than $\mathrm{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ because the estimation of $\mathrm{TE}_\mu(\boldsymbol{\gamma}, \boldsymbol{\psi})$ is unaffected by any dependence between ϵ and the other inputs.

Above we highlighted that $\mathrm{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ can be suppressed by dependence between ϵ and the other inputs. Since the total variation in Y can be decomposed into

FOE($\boldsymbol{X'}, \epsilon$) and TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) by the law of total variance, any suppression of TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) corresponds to an inflation of FOE($\boldsymbol{X'}, \epsilon$). This inflation is because dependence between ϵ and other inputs results in the individual effect of each input not being fully disentangled, and so variation that is really due to $\boldsymbol{\gamma}$ or $\boldsymbol{\psi}$ is included in FOE($\boldsymbol{X'}, \epsilon$). However as shown previously, such dependence does not affect the decomposition of the total variation in Y into FOE(μ) = var(μ) and TE(ϵ) = var(ϵ). Therefore estimating var(μ) and then TE$_\mu$($\boldsymbol{\gamma}, \boldsymbol{\psi}$) avoids the effect of the dependence between ϵ and other inputs, and so allows the effect of $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$ to be disentangled from the effect of ϵ. As a result, for the remainder of the thesis we shall suppress the subscript μ on total effects such as TE$_\mu$($\boldsymbol{\gamma}, \boldsymbol{\psi}$), and refer to total effects on μ simply as total effects.

With respect to the interpretation of the terms in Equation (2.11), the combined total effect (TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$)) of $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$ is included. The predictor means $\boldsymbol{\gamma}$ and random effects $\boldsymbol{\psi}$ are the inputs that capture variation across the different ensemble members, and so their total effect is the explained variation in precipitation that is found to be due to the differences between ensemble members. As discussed in the previous paragraph, however, this can be interpreted as the total variation in *precipitation* that is found to be due to the differences between the ensemble members. Relating TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) to the representation of the physical process captured by the GLMM, a more formal interpretation of TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) is

Interpretation 2.3. the total explained variation in precipitation due to the variation across ensemble members in the representation of the physical process and precipitation bias.

This is a measure of the extent to which the physical process captured by our GLMM is represented robustly across the climate models.

The other term in Equation (2.11) is the first-order effect (FOE($\boldsymbol{X'}$)) of $\boldsymbol{X'}$, or

Interpretation 2.4. the variation in precipitation solely due to the temporal variation in the represented atmospheric conditions.

This is a measure of how prominent the represented physical process is in driving the precipitation within each ensemble member. The terms in Equation (2.11) will often be more informative when normalized by the total variation in the precipitation. At this stage in our decomposition we can use Equation (2.11) and Interpretations 2.3 and 2.4 to develop the diagram of Figure 2.1 into Figure 2.2.

Both of the terms in Equation (2.11) are informative, but it is estimating the total effect (TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$)) of the predictor means and random effects (Interpretation 2.3) in particular that could be the objective of an analysis. Recall from Section 1.7 that

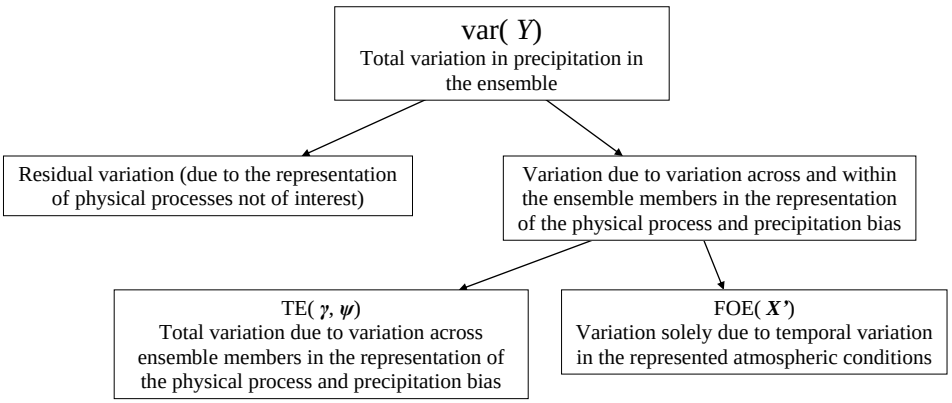

Figure 2.2: First two parts of the proposed decomposition structure for precipitation (Y) in an ensemble of climate models.

a random-effects ANOVA can be used to estimate the variation across ensemble members in the distribution (mean) of the precipitation. Therefore, comparing $\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ to this result from a random-effects ANOVA will show the additional variation that is found to be due to the differences between the climate models when predictors are included in the statistical modelling to capture the variation in the representation of the physical process of interest.

Further detail can be gained from $\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$. This is because $\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ can be decomposed into the first-order effect ($\text{FOE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$) of the same inputs, and the interactions between these inputs ($\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$) and the temporal variation $\boldsymbol{X'}$ in the predictors, i.e.

$$\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi}) = \text{FOE}(\boldsymbol{\gamma}, \boldsymbol{\psi}) + \text{int}(\boldsymbol{X'}, \boldsymbol{\gamma}) + \text{int}(\boldsymbol{X'}, \boldsymbol{\psi}) + \text{int}(\boldsymbol{X'}, \boldsymbol{\gamma}, \boldsymbol{\psi}) \ , \tag{2.13}$$

where, recalling Equation (2.12), the first-order effects and interactions are the terms that would be obtained whether decomposing $\text{var}(Y)$ or $\text{var}(\mu)$. Since $\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ is the total variation found to be due to the variation across ensemble members in the representation of the physical process and precipitation bias (recall Interpretation 2.3), $\text{FOE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ is the variation solely due to these sources. Given this, the diagram of Figure 2.2 can be developed into Figure 2.3.

Before discussing the interactions in Equation (2.13), the first-order effect ($\text{FOE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$) of the predictor means and random effects can be decomposed into the individual first-order effect ($\text{FOE}(\boldsymbol{\gamma})$) of the predictor means, that ($\text{FOE}(\boldsymbol{\psi})$) of random effects, and the interaction ($\text{int}(\boldsymbol{\gamma}, \boldsymbol{\psi})$) between them. As such, Equation (2.13) becomes

$$\begin{aligned}\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi}) = \text{FOE}(\boldsymbol{\gamma}) + \text{FOE}(\boldsymbol{\psi}) + \text{int}(\boldsymbol{X'}, \boldsymbol{\gamma}) + \text{int}(\boldsymbol{X'}, \boldsymbol{\psi}) + \text{int}(\boldsymbol{\gamma}, \boldsymbol{\psi}) \\ + \text{int}(\boldsymbol{X'}, \boldsymbol{\gamma}, \boldsymbol{\psi}) \ .\end{aligned} \tag{2.14}$$

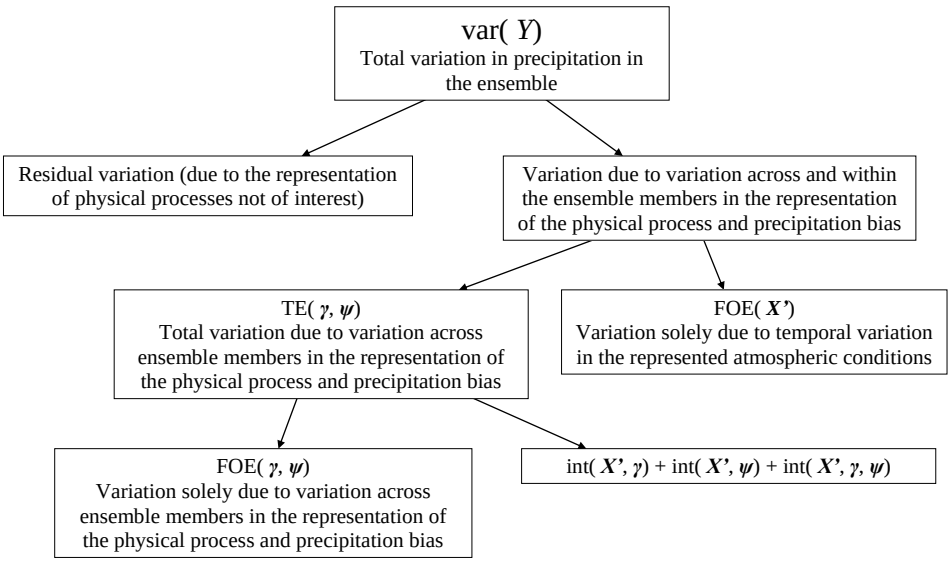

Figure 2.3: First three parts of the proposed decomposition structure for precipitation (Y) in an ensemble of climate models. Total effects (TE) (on μ), first-order effects (FOE) and interactions (int) are used for the temporal components of the predictors ($\boldsymbol{X'}$), the predictor means ($\boldsymbol{\gamma}$), and the random effects ($\boldsymbol{\psi}$).

There are many terms in Equation (2.14), but this is a general decomposition and the presence of all of the interactions depends on the specification of the GLMM. For example, if the predictor means ($\boldsymbol{\gamma}$) do not interact with the temporal components ($\boldsymbol{X'}$) of the predictors, then only FOE($\boldsymbol{\gamma}$), FOE($\boldsymbol{\psi}$), int($\boldsymbol{X'}, \boldsymbol{\psi}$) and int($\boldsymbol{\gamma}, \boldsymbol{\psi}$) would remain from Equation (2.14). Moreover, returning to the general decomposition, Equation (2.14) can be simplified by grouping the terms involving the predictor means $\boldsymbol{\gamma}$ into their total effect, i.e.

$$\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi}) = \text{FOE}(\boldsymbol{\psi}) + [\text{int}(\boldsymbol{X'}, \boldsymbol{\psi}) + \text{TE}(\boldsymbol{\gamma})] \ . \tag{2.15}$$

The first term of Equation (2.15) is the first-order effect (FOE($\boldsymbol{\psi}$)) of the random effects. As can be seen from its mathematical definition, $\text{var}_{\boldsymbol{\psi}}[E_{\mu}(\mu \mid \boldsymbol{\psi})]$, this is the variation across ensemble members in precipitation bias (i.e. mean precipitation rate) remaining after averaging over the total variation ($\boldsymbol{\gamma}$ and $\boldsymbol{X'}$) in the predictors. Given that the predictors capture the represented atmospheric conditions of the physical process of interest, FOE($\boldsymbol{\psi}$) is

Interpretation 2.5. the variation across ensemble members in precipitation bias remaining after averaging over the represented atmospheric conditions.

As discussed in Section 2.1, the source of this is partly the variation across ensemble members in the relationships between the predictors and precipitation, and partly additional variation across the members in mean precipitation rate (precipitation bias). This precipitation bias could be due to the representation of any physical

process, and includes any precipitation bias independent of the physical process of interest.

Aside from FOE($\boldsymbol{\psi}$), Equation (2.15) involves the interaction (int($\boldsymbol{X'}, \boldsymbol{\psi}$)) between the temporal components of the predictors and the random effects, and the total effect (TE($\boldsymbol{\gamma}$)) of the predictor means. If FOE($\boldsymbol{\psi}$) includes the precipitation bias independent of the physical process of interest, then the source of int($\boldsymbol{X'}, \boldsymbol{\psi}$) and TE($\boldsymbol{\gamma}$) is variation across ensemble members in the representation of the physical process of interest. As such, summing int($\boldsymbol{X'}, \boldsymbol{\psi}$) and TE($\boldsymbol{\gamma}$) gives

Interpretation 2.6. the minimum amount of explained variation in precipitation found to be due to the variation across ensemble members in the representation of the physical process.

This is the *minimum* amount of variation found to be due to this source because variation across ensemble members in the representation of the physical process of interest can also contribute to FOE($\boldsymbol{\psi}$) (as discussed above).

We can use Equations (2.14) and (2.15), and Interpretations 2.5 and 2.6, to develop the diagram of Figure 2.3 into Figure 2.4.

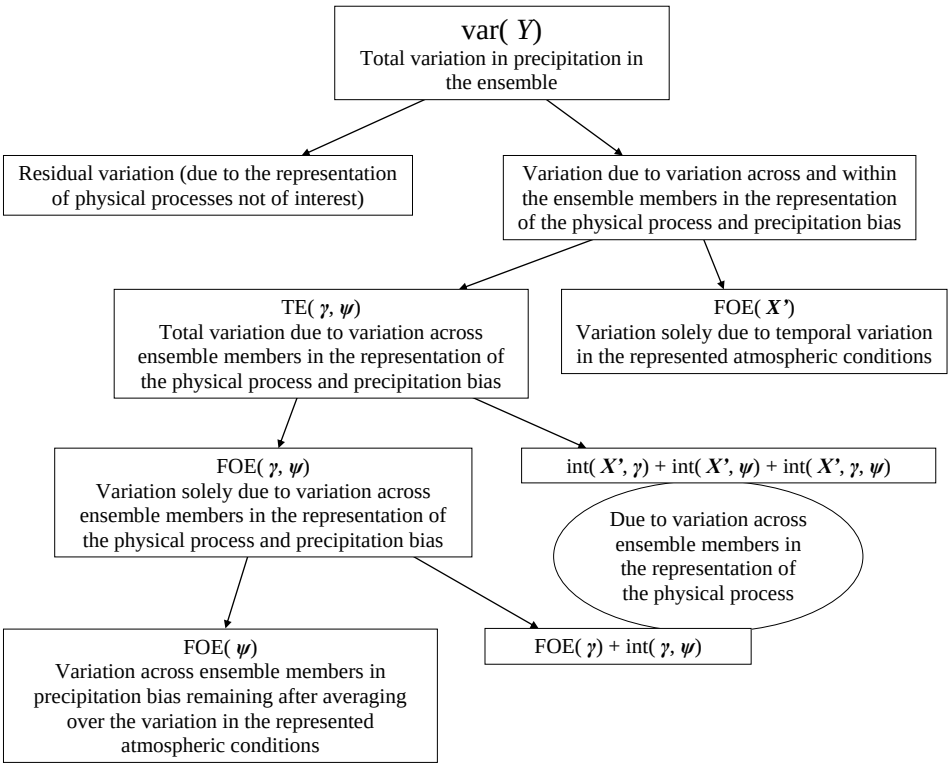

Figure 2.4: First four parts of the proposed decomposition structure for precipitation (Y) in an ensemble of climate models. Total effects (TE) (on μ), first-order effects (FOE) and interactions (int) are used for the temporal components of the predictors ($\boldsymbol{X'}$), the predictor means ($\boldsymbol{\gamma}$), and the random effects ($\boldsymbol{\psi}$).

In addition to quantifying the minimum amount of variation in precipitation found to be due to the variation across ensemble members in the representation of the physical process, detail can be gained into why the representation of the physical process varies across climate models. This is because variation across ensemble members in the representation of the physical process can be because 1) the relationship between the represented atmospheric conditions and precipitation varies across the members, 2) the frequency distribution of the represented atmospheric conditions varies across the members, or 3) a combination of (1) and (2). In the case of (1), given atmospheric conditions would lead to different rates of precipitation in different ensemble members, whereas in the case of (2), given atmospheric conditions would be simulated with different frequencies in different ensemble members.

Examining the terms in Equation (2.14), the interaction ($\text{int}(\boldsymbol{X'}, \boldsymbol{\psi})$) between the temporal components of the predictors and the random effects is

Interpretation 2.7. the additional variation in precipitation due to the joint variation in the temporal component of the represented atmospheric conditions and its relationship with precipitation, unaccounted for by their individual first-order effects.

As such, $\text{int}(\boldsymbol{X'}, \boldsymbol{\psi})$ is driven by variation across ensemble members in the relationship between the represented atmospheric conditions and the precipitation ((1) above). This differs from the first-order effect ($\text{FOE}(\boldsymbol{\gamma})$) of the predictor means, which is

Interpretation 2.8. the variation in precipitation solely due to the variation across ensemble members in the frequency distribution of the represented atmospheric conditions.

As such, $\text{FOE}(\boldsymbol{\gamma})$ is driven by (2) above. The other terms in Equation (2.14) are driven by a combination ((3) above) of variation across ensemble members in the relationship between the represented atmospheric conditions and the precipitation, and the variation across the members in the frequency distribution of the represented atmospheric conditions.

The implication of $\text{int}(\boldsymbol{X'}, \boldsymbol{\psi})$ being the only term in Equation (2.14) that is driven only by (1) above is that $\text{int}(\boldsymbol{X'}, \boldsymbol{\psi})$ can be used to gain insight into the extent to which the relationship between the represented atmospheric conditions and the precipitation varies across ensemble members. Similarly, $\text{FOE}(\boldsymbol{\gamma})$ can be used to gain insight into the extent to which the frequency distribution of the represented atmospheric conditions varies across ensemble members ((2) above). For example suppose that $\text{int}(\boldsymbol{X'}, \boldsymbol{\psi})$ dominated the variation in precipitation that was found to be due to the variation across ensemble members in the representation of a physical

process. This would imply that the representation of the physical process varies across ensemble members because given atmospheric conditions lead to different rates of precipitation in different ensemble members, rather than because given atmospheric conditions are simulated with different frequencies in different ensemble members.

As seen previously in Section 2.1, the first-order effect (FOE($\boldsymbol{\psi}$)) of the random effects can include variation across ensemble members in precipitation bias due to the representation of any physical process, including the physical process of interest. Moreover it will often not be useful to decompose FOE($\boldsymbol{\psi}$) into the individual contribution of each random effect. Therefore FOE($\boldsymbol{\psi}$) includes the variation in mean precipitation rate after correcting for the process of interest, and so although FOE($\boldsymbol{\psi}$) quantifies variation due to the differences between the ensemble members, beyond this its cause is generally unknown (i.e. not necessarily due to the process of interest in an application). An implication of this is that the total explained variation in precipitation due to the variation across ensemble members in the relationship between the represented atmospheric conditions and the precipitation cannot be fully quantified. This is because although part of this variation can be quantified by the interactions involving the random effects seen previously in Equation (2.14) (e.g. int($\boldsymbol{X'}, \boldsymbol{\psi}$), see Interpretation 2.7), the variation in the relationship can also contribute to FOE($\boldsymbol{\psi}$).

Despite interactions such as int($\boldsymbol{X'}, \boldsymbol{\psi}$) not quantifying the total explained variation in precipitation due to the variation across ensemble members in the relationship between the represented atmospheric conditions and the precipitation, it is possible that such interactions would quantify much of it. This is because the extent to which the differences between climate models affect their simulation of precipitation could depend on the represented atmospheric conditions. For example a physical process could be represented robustly across climate models in average atmospheric conditions when perhaps heavy precipitation is unlikely, but could be less robustly represented when extreme atmospheric conditions are represented and perhaps heavy precipitation is likely (see the left panel of Figure 2.5 for a schematic).

The left panel of Figure 2.5 shows a scenario when a physical process is represented robustly across climate models in average atmospheric conditions but is represented less robustly in extreme atmospheric conditions. In this scenario, variation would be required in the represented atmospheric conditions for the variation across ensemble members in the relationship between the represented atmospheric conditions and the precipitation to account for much variation in the precipitation. The result would be that the total variation in the relationship would contribute little to FOE($\boldsymbol{\psi}$), and so

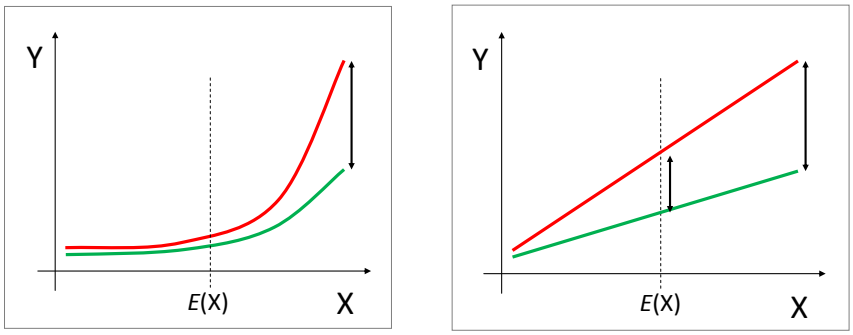

Figure 2.5: Schematic showing that the extent to which the variation in precipitation (Y) due to variation across the members of an ensemble in the relationship between predictor X and the precipitation can depend on the variation in X. This is due to interaction between X and its relationship with Y. In both panels there is greatest variation in Y due to the variation across ensemble members in the relationship at high X, and there is little variation in Y due to the variation in the relationship at average (or low) X in the left panel.

the interactions involving the random effects in Equation (2.14) would approximate the robustness of the relationship well. Indeed, if these interactions contribute little to the variation in the precipitation, then this often implies that the relationship is robust. This is the case with linear relationships, for example, because any variation across ensemble members in the relationship would contribute the most variation to the precipitation when extreme atmospheric conditions are represented (see the right panel of Figure 2.5 for a schematic).

Up to this point we have grouped all predictor means ($\boldsymbol{\gamma}$) together, and grouped all temporal components ($\boldsymbol{X'}$) of the predictors together. Unlike with the random effects, estimating importance measures of the within-model mean and temporal component of individual predictors, and their interactions with each other and the random effects, can help identify the dominant features of the process. This could provide further insight into why the process varies across and within the ensemble members. For example it could be that multiple predictors are used, but that one predictor dominates the process and the other predictors have little influence on precipitation. Highlighting this detail could be done by estimating importance measures of the individual inputs and their interactions.

Depending on the GLMM, it could be that it is desirable to always consider some inputs grouped together when estimating importance measures, treating them as a single input. This could be the case if transformations of a predictor are included (e.g. a quadratic and cubic transformation) to capture a non-linear relationship. To illustrate this, consider an extension to the GLMM of Section 1.9,

$$
\begin{aligned}
Y_{ij} \mid \left(\psi_j^{(0)}, \psi_j^{(1)}, \psi_j^{(2)}, \psi_j^{(3)}, X_{ij}^{(1)}, X_{ij}^{(2)}\right) &\sim \text{Gamma}(\mu_{ij}, \lambda) \\
\log(\mu_{ij}) &= \beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \psi_j^{(0)} + \psi_j^{(1)} X_{ij}^{(1)} + \psi_j^{(2)} X_{ij}^{(2)} \\
&\quad + \beta_3 (X^{(1)})_{ij}^2 + \psi_j^{(3)} (X^{(1)})_{ij}^2 \\
\left(\psi_j^{(0)}, \psi_j^{(1)}, \psi_j^{(2)}, \psi_j^{(3)}\right) &\sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}) \ ,
\end{aligned}
$$

where a quadratic transformation of predictor $X^{(1)}$ is added, along with an associated random effect ($\psi^{(3)}$). In this case it may be of interest to split up the two components of the relationship between $X^{(1)}$ and Y, and consider (for example) $\text{int}(X'^{(1)}, \psi^{(1)})$ and $\text{int}(X'^{(1)}, \psi^{(3)})$ separately. However it will often be desirable to group $\psi^{(1)}$ and $\psi^{(3)}$ together and treat them as one input. As already seen, grouping inputs is trivial with importance measures, and therefore the methodology presented in this chapter can account for non-linear relationships between predictors and the response of the GLMM.

Yip et al. (2011) provide a framework for partitioning the uncertainty in an ANOVA statistical model fitted to data from an ensemble of climate models into what they call 'climate model uncertainty' and 'internal variability' (and emission scenario uncertainty, but we will focus on the other two sources), and their work is referenced in Section 1.9. Despite their focus being on ANOVAs instead of GLMMs, we can end this subsection by comparing the interpretations of some of the terms of our general decomposition with those of Yip et al. (2011). Equation (2.11) on page 53 decomposes the variation in precipitation into the total effect ($\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$) of the predictor means and random effects, the first-order effect ($\text{FOE}(\boldsymbol{X'})$) of the temporal variation in the predictors, and the remaining residual variation ($\text{var}(\epsilon)$), and Equation (2.13) on page 56 decomposes $\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ into the first-order effect ($\text{FOE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$) of the same predictors and interactions between these inputs ($\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$) and the temporal variation $\boldsymbol{X'}$ in the predictors. Comparing the terms to those of Yip et al. (2011), their climate model uncertainty term is equivalent to $\text{FOE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$, whereas their internal variability term is included in the sum of $\text{FOE}(\boldsymbol{X'})$ and $\text{var}(\epsilon)$. In addition to $\text{FOE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$, $\text{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$ in our decomposition accounts for the interaction between climate model uncertainty and internal variability, which we interpret as the variation due to climate model uncertainty that also depends on internal variability (note that Yip et al., 2011, interpret what they call 'scenario-dependent climate model uncertainty', as opposed to scenario-independent climate model uncertainty, in a similar way, but with respect to interaction with emission scenario instead of the internal variability).

2.2.2 Dependence between inputs

When using a GLMM to analyse the representation of a driving physical process of precipitation by climate models, the inputs are 1) the temporal component of each predictor within the climate models, 2) the mean within each climate model of each predictor, and 3) each random effect. The inputs were independent in the analytical example of Section 2.1, and independence between inputs was assumed when interpreting the terms of the general decomposition outlined in the previous subsection. Some inputs will, however, often be dependent. We will discuss different options for modelling the inputs and accounting for dependence between them later, but we discuss the possible impact of dependence between inputs in this subsection.

The temporal variation in the different predictors within a climate model ((1) above) will often be dependent because the predictors capture the atmospheric conditions represented in the climate model. Moreover, the variation across different climate models in the mean of each predictor and that in each random effect ((2) and (3) above) will often be dependent because they share the same source: the differences between the climate models. Therefore when specifying a model for the inputs, we propose that general guidance should be to account for dependence between the temporal variation in different predictors, and between the predictor means and random effects. However we also propose that independence can often be assumed between the inputs that capture the temporal variation within each ensemble member, and those that capture variation due to the differences between the ensemble members. Indeed, the inputs that capture the temporal variation in the predictors are centred on zero within each ensemble member by their definition, and so the first-order dependence between these inputs and those that capture variation due to the differences between the ensemble members will always be zero. Moreover even if this assumption is found not to be valid for a physical process, then making the assumption may lead to results that are sufficient approximations for a given application. As a result, we recommend that this independence assumption is made unless reason is found that it is not valid for any given application, but that dependence is accounted for between the predictor means and random effects, and between the temporal variation in the predictors.

Assuming independence between the temporal variation $\boldsymbol{X'}$ in the predictors and the other inputs ($\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$) has the advantage that certain terms are orthogonal when decomposing the total explained variation in precipitation. In particular, Equation 2.11 (page 53) decomposed the variation accounted for by the GLMM into the total variation due to the differences between ensemble members ($\mathrm{TE}(\boldsymbol{\gamma}, \boldsymbol{\psi})$) and the variation solely due to the temporal variation in the represented atmospheric

conditions within each member (FOE($\boldsymbol{X'}$)). When $\boldsymbol{X'}$ is independent of $\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$, the interpretations of FOE($\boldsymbol{X'}$) and TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) are unaffected by dependence, and so the total variation due to the differences between ensemble members can be fully disentangled from the temporal variation in the represented atmospheric conditions of interest.

Figure 2.6 summarizes our proposed structure for decomposing the variation in precipitation in an ensemble of climate models (although further detail can be gained by separating the effects of each predictor, separating interactions, etc). The figure develops Figure 2.3 by using dashed lines to indicate where the interpretations of the terms are affected by dependence when the dependence structure recommended above is used. As highlighted by Figure 2.6, using the dependence structure recommended above means that dependence between inputs could affect the decomposition when disentangling the variation across the climate models in precipitation bias (FOE($\boldsymbol{\psi}$)) from part of the variation across the models in the representation of the physical process (FOE($\boldsymbol{\gamma}$) + int($\boldsymbol{\gamma}, \boldsymbol{\psi}$)). This would be due to dependence between the random effects and the predictor means. For example it could be that a strong relationship between a predictor and precipitation tends to be compensated for by a low mean of that predictor. Dependence between random effects and predictor means implies that the causes of variation in precipitation that are found to be due to the differences between the ensemble members cannot be fully disentangled.

Even in the presence of dependence, however, some detail can usually be gained when disentangling the effects of dependent inputs. This is because the dependence may be weak, and the mathematical definition and formal interpretation of a first-order effect, those of a total effect and those of an interaction are meaningful in themselves (see Interpretations 1.1, 1.2 and 1.3, from page 32). As a result, estimating the terms shown near the bottom of Figure 2.6 will often remain informative in the presence of dependence. This is also the case when attempting to disentangle the effect of each predictor and each interaction from the others when multiple predictors are present. For example we can consider disentangling the first-order effect (FOE($\boldsymbol{X'}$)) of the temporal variation in the predictors ($X^{(1)}$ and $X^{(2)}$) into

$$\text{FOE}(X'^{(1)}) + \text{FOE}(X'^{(2)}) + \text{int}(X'^{(1)}, X'^{(2)}) \, .$$

If $X'^{(1)}$ and $X'^{(2)}$ do not interact, and their individual first-order effects sum to approximately their combined first-order effect, then their dependence has not substantially affected the estimation of their individual first-order effects. Therefore we propose that as long as the presence of dependence between inputs is made clear, and the importance measures are clearly defined, dependence generally should not prevent the effects of individual inputs from being estimated.

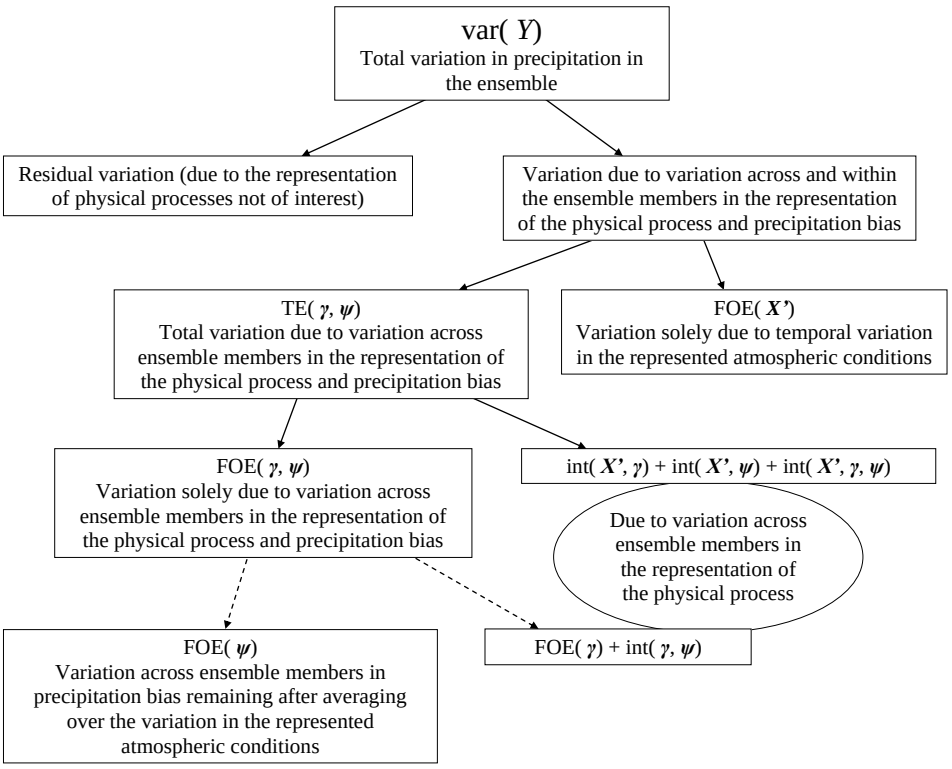

Figure 2.6: Decomposition structure for precipitation (Y) in an ensemble of climate models. Total effects (TE), first-order effects (FOE) and interactions (int) are used for the temporal components of the predictors ($\boldsymbol{X'}$), the predictor means ($\boldsymbol{\gamma}$), and the random effects ($\boldsymbol{\psi}$). Dashed lines indicate when non-orthogonality can occur due to dependence between $\boldsymbol{\psi}$ and $\boldsymbol{\gamma}$.

2.3 Numerical example

To demonstrate using simulation to estimate importance measures, suppose the following GLMM is used:

$$
\begin{aligned}
Y_{ij} &= 3 + 6X_{ij}^{(1)} + 2X_{ij}^{(2)} + \psi_j^{(0)} + \psi_j^{(1)}X_{ij}^{(1)} + \psi_j^{(2)}X_{ij}^{(2)} + \epsilon_{ij} \\
X_{ij}^{(1)} &= \gamma_j^{(1)} + X_{ij}^{\prime(1)} \qquad \gamma_j^{(1)} \sim N(1, 36) \\
\psi_j^{(0)} &\sim N(0, 25) \qquad X_{ij}^{\prime(1)} \sim N(0, 9) \\
\psi_j^{(1)} &\sim N(0, 16) \qquad X_{ij}^{(2)} \sim N(1, 4) \\
\psi_j^{(2)} &\sim N(0, 9) \qquad \epsilon_{ij} \sim N(0, 100) \ ,
\end{aligned}
$$

where i indexes time and j indexes the member of an ensemble of climate models. If, for example, there were 2000 time points and five ensemble members, then i would take the values 1,...,2000 and j would take 1,...,5. The random effects are $\psi^{(0)}$, $\psi^{(1)}$ and $\psi^{(2)}$, the predictors are $X^{(1)}$ and $X^{(2)}$, the residual term is ϵ, and all inputs are independent. Predictor $X^{(1)}$ is decomposed into its mean ($\gamma^{(1)}$) within each ensemble member and its temporal deviations ($X'^{(1)}$) from these means. The distribution of predictor $X^{(2)}$ does not vary across the ensemble members. This GLMM extends

that used in the Section 2.1 by an additional predictor ($X^{(2)}$) and random effect ($\psi^{(2)}$), but remains simple enough to compute the importance measures analytically. The advantage of analytical computations with this example is that the importance measures estimated by simulation can be compared against their true values. The presence of an additional predictor and random effect allows their comparison with $X^{(1)}$ and $\psi^{(1)}$.

With respect to notation when using simulation to estimate importance measures, we can follow the notation used previously in Subsection 1.10.3 (also recall the simulation algorithm outlined in Subsection 1.10.3, from page 40). As such, to simulate (e.g.) the first-order effect ($\text{var}_{X'^{(1)}}[E_Y(Y \mid X'^{(1)})]$) of $X'^{(1)}$, let R be the number of samples taken for $X'^{(1)}$, and for each of these let r be the number of samples taken for each of the other inputs (such that rR samples are taken in total for each input other than $X'^{(1)}$). Similarly, to simulate the total effect ($E_{-X'^{(1)}}[\text{var}_Y(Y \mid -X'^{(1)})]$) of $X'^{(1)}$, rR samples are taken in total for $X'^{(1)}$ and R samples taken for each of the other inputs.

The black values in Figure 2.7 show the analytical decomposition of the variation in Y. The sources of variation are expressed as percentages of the total variation in Y, rounded to zero decimal places (this rounding is the reason that $\text{TE}(\gamma^{(1)}, \psi^{(0)}, \psi^{(1)}, \psi^{(2)})$, $\text{FOE}(X'^{(1)})$ and $\text{FOE}(X^{(2)})$ do not sum to the figure reported for the total variation accounted for by the GLMM). Of the total variation (2542) in Y, 96% is accounted for by the GLMM. Only 13% of the variation in Y is solely due to the temporal variation in $X^{(1)}$, and 1% solely to the variation in $X^{(2)}$. This is because the majority of the variation (83%) in Y is due to the differences between the ensemble members. The largest source of this is variation across ensemble members in the frequency distribution of $X^{(1)}$, although there is also variation across the members in the relationship between $X^{(1)}$ and Y. This can be seen because the first-order effect of $\gamma^{(1)}$ is 51% and its interaction with $\psi^{(1)}$ is 23%, whereas the interactions between the temporal variation in $X^{(1)}$ and its relationship with Y is smaller (6%). The variation due to the interaction between $X^{(2)}$ and its relationship with Y is small (1%), as is the variation across ensemble members in the bias in Y remaining after averaging over the variation in the predictors ($\text{FOE}(\psi^{(0)}, \psi^{(1)}, \psi^{(2)}) = 2\%$). In summary, the variation in $X^{(2)}$ is unimportant relative to that in $X^{(1)}$, with considerable variation across ensemble members in the distribution of $X^{(1)}$, some variation across the members in its relationship with Y, and some temporal variation.

The red values of Figure 2.7 show the decomposition of the variation in Y as estimated using simulation from the true model. Comparing this decomposition against

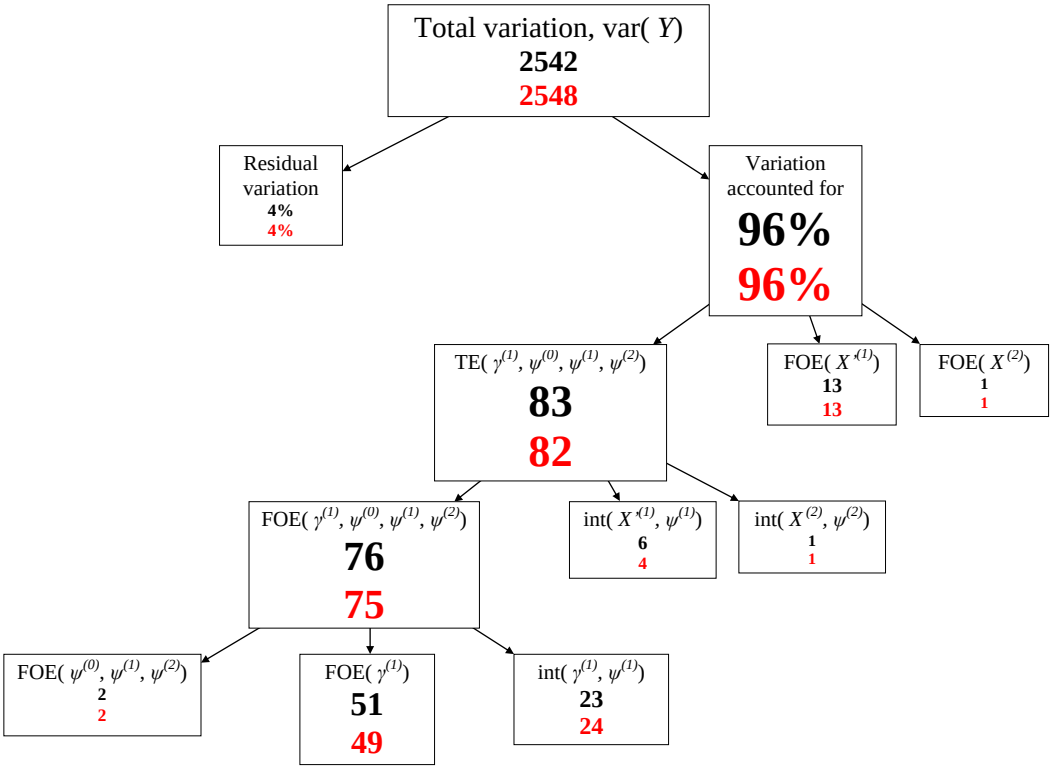

Figure 2.7: Global sensitivity analysis results for the numerical example, showing total effects (TE), first-order effects (FOE) and interactions (int), all expressed as percentages of the total variation in Y. Black values are the true values; red values show simulated estimates based on $R = 1000$ and $r = 2000$ samples from the true model. All values are rounded to zero decimal places.

the true values, each simulated estimate has an error of no more than 2% of the total variation in Y (along with the rounding, this sampling error causes the inconsistency seen in parts of the decomposition, e.g. FOE$(\psi^{(0)}, \psi^{(1)}, \psi^{(2)})$, FOE$(\gamma^{(1)})$ and int$(\gamma^{(1)}, \psi^{(1)})$ do not sum to FOE$(\gamma^{(1)}, \psi^{(0)}, \psi^{(1)}, \psi^{(2)})$ with the values shown). For example the first-order effect of $\gamma^{(1)}$ is estimated to be 49% when using simulation here, instead of as its true value of 51%. Although this relative percentage error is only 4%, the relative percentage error of the simulated estimate of int$(X'^{(1)}, \psi^{(1)})$ from its true value is large (33%). This highlights that interactions can be more difficult to estimate than other importance measures because they require manipulating multiple importance measures (recall Equation (1.5) on page 34), each of which could have sampling error.

The accuracy of the estimates depends on the number of samples used. Consider (for example) the first-order effect $(\mathrm{var}_{X'^{(1)}}[E_Y(Y \mid X'^{(1)})])$ of $X'^{(1)}$. To simulate this term, a value of $X'^{(1)}$ is simulated, r values of each of the other inputs are simulated, this is repeated R times, and these samples are used to compute the importance measure. The simulated estimates shown by the red values in Figure 2.7 are based on $R = 1000$ samples of size $r = 2000$ (which took only a few seconds to run the whole decomposition). Increasing the sample size of the simulation will tend to increase the accuracy of the decomposition, see Table 2.1 for a sensitivity

analysis to the number of samples used to estimate the first-order effect of $X'^{(1)}$. It can be seen from Table 2.1 that both R and r are important, but that accurate results are obtained when R and r are at least 800 in this example. Table 2.1 also highlights that increasing the number of samples will not always increase the accuracy of the estimates, but the sampling error will be small if enough samples are used. The number of samples required depends on the accuracy required, and different statistical models will require a different number of samples to achieve a given level of accuracy. The sensitivity of simulated importance measures to the number of samples used can be tested by simulating additional samples and examining the extent to which the importance measures change. This is done in each of the applications in the later chapters to decide on the number of samples used, such that the results of no application would particularly change if more samples were taken.

	$r = 100$	200	400	800	1600	3200	6400	12800
$R = 100$	7	14	16	18	20	19	21	21
200	7	3	2	3	4	4	5	5
400	11	5	5	5	3	3	2	2
800	5	2	1	1	0	0	0	0
1600	5	3	2	1	1	1	1	1
3200	5	2	0	1	1	1	1	1
6400	7	3	1	1	0	0	0	0
12800	7	3	1	0	0	0	0	0

Table 2.1: Sensitivity analysis to the number of samples (R and r) used to estimate the first-order effect ($\text{var}_{X'^{(1)}}[E_\mu(\mu \mid X'^{(1)})]$) of input $X'^{(1)}$. The relative percentage error (to zero decimal places) from the true value of 324 is shown.

The samples used to simulate the total effect of an input can be examined to provide detail of how its interaction with other inputs affects the variation in Y. To see this, consider simulating an estimate for the total effect ($E_{-\psi^{(1)}}[\text{var}_\mu(\mu \mid -\psi^{(1)})]$) of $\psi^{(1)}$ (where μ is the expectation of Y). This requires holding each of the inputs except for $\psi^{(1)}$ constant, estimating the total variation ($\text{var}_\mu(\mu \mid -\psi^{(1)})$) in the expectation of Y due to the variation in $\psi^{(1)}$, repeating this and averaging over the distributions of the other inputs. Before averaging over the distributions of the other inputs, however, information can be gained from examining the samples. Figure 2.8 shows how $\text{var}_\mu(\mu \mid -\psi^{(1)})$ depends on selected inputs. If there is a relationship between $\text{var}_\mu(\mu \mid -\psi^{(1)})$ and an input, then the importance of $\psi^{(1)}$ depends on the variation in that input because they interact. This is the case with $\gamma^{(1)}$ and with $X'^{(1)}$, see the bottom panels of Figure 2.8. How each interaction affects the variation in Y can also be seen. It is clear that the variation in $\psi^{(1)}$ influences Y the most at high or low $\gamma^{(1)}$, and to a lesser extent at high or low $X'^{(1)}$. The importance measures

can estimate the extent to which each interaction influences Y, and bivariate plots like those of Figure 2.8 do not highlight the number of inputs involved in each interaction (e.g. a three-dimensional plot would be required to fully show the effect of a three-way interaction), but examining the samples can provide further detail to the importance measures.

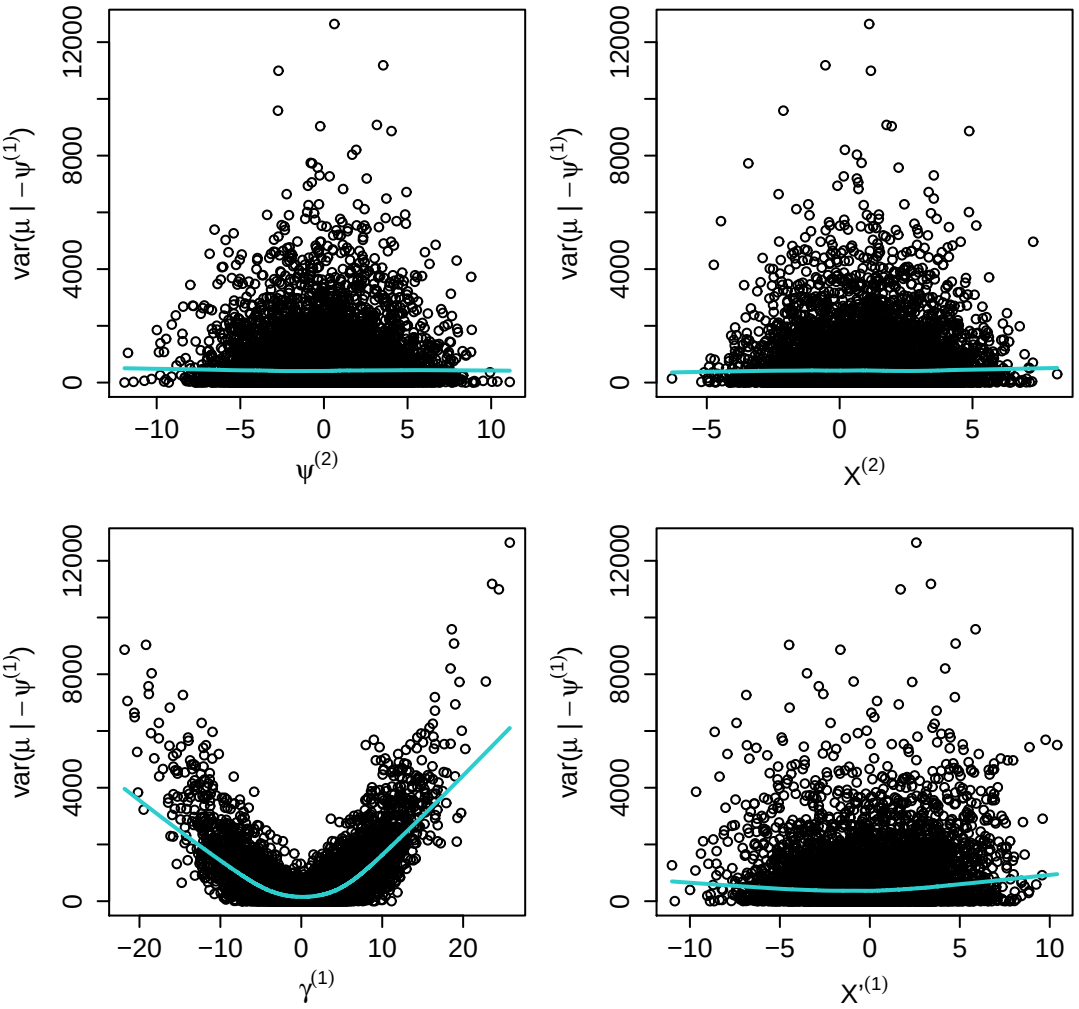

Figure 2.8: Estimated total variation ($\text{var}_\mu(\mu \mid -\psi^{(1)})$) in the expectation μ of Y due to the variation in $\psi^{(1)}$, plotted against simulated values of selected inputs. Curves are local polynomial smoothers. Taking the mean of $\text{var}_\mu(\mu \mid -\psi^{(1)})$ provides an estimate for the total effect of $\psi^{(1)}$, $E_{-\psi^{(1)}}[\text{var}_\mu(\mu \mid -\psi^{(1)})]$.

2.3.1 Visualizing importance measures

To demonstrate how the importance measures can be visualized (i.e. seen in the data), a synthetic data set is simulated from the model above. Five values for each random effect and mean of $X^{(1)}$ are simulated (i.e. five ensemble members), and two thousand values for $X'^{(1)}$ and $X^{(2)}$ for each of these ensemble members (i.e. two thousand temporal points), such that $j = 1, ..., 5$ and $i = 1, ..., 2000$ in the synthetic

data set. The total number of data points in the synthetic data set is therefore 10 000.

The decomposition of Figure 2.7 showed that the dominant sources of variation in Y are the variation across ensemble members in the distribution (mean) of $X^{(1)}$, the temporal variation in $X^{(1)}$, and the variation across ensemble members in the relationship between $X^{(1)}$ and Y. The variation in $X^{(2)}$, and the variation across ensemble members in its relationship with Y, are relatively unimportant. This can be visualized from the synthetic data, which are displayed in Figure 2.9. The top-left panel of Figure 2.9 shows that there is substantial variation across ensemble members in the distribution of Y, relative to its temporal variation. From the top-right panel of Figure 2.9 it can be seen that there is also substantial variation across ensemble members in the distribution of $X^{(1)}$, and this variation positively correlates with the variation across the members in Y. No variation across ensemble members is seen in the distribution of $X^{(2)}$ (no box plots shown).

The bottom panels of Figure 2.9 show that $X^{(1)}$ has a greater impact on Y than $X^{(2)}$. This can be seen because the relationship is relatively strong between $X^{(1)}$ and Y in all levels except for b, whereas there is a weaker relationship between $X^{(2)}$ and Y in all empirical levels other than c. Furthermore the empirical points are tighter to the curves (bottom-left panel) with $X^{(1)}$, and there is greater variation in $X^{(1)}$ than $X^{(2)}$. With respect to the variation across ensemble members in the relationships, there appears similar absolute variation in the relationship between $X^{(1)}$ and Y, and in the relationship between $X^{(2)}$ and Y, suggesting that the variance components of $\psi^{(1)}$ and $\psi^{(2)}$ are similar. However this does not necessarily lead to a similar impact on Y of course, which is why we require importance measures.

It is clear that the strong positive correlation between the distributions of $X^{(1)}$ and Y seen in the box plots of the top panels of Figure 2.9 is weakened by the negative relationship between $X^{(1)}$ and Y in level b, and by the positive relationship between $X^{(2)}$ and Y in level c. However the mean of $X^{(1)}$ is lowest in level a, yet the mean of Y in this level is greater than it is in level e. There appears no reason involving the predictors for this, which suggests that there is greater bias in Y in level a than in level e due to higher $\psi^{(0)}$ in level a.

Overall, Figure 2.9 can be used in this example to determine that the main drivers of the variation in Y are the variation across ensemble members in the distribution of $X^{(1)}$, its temporal variation within each level, and the variation in its relationship with Y. It is these conclusions for which the importance measures provide a formal quantification.

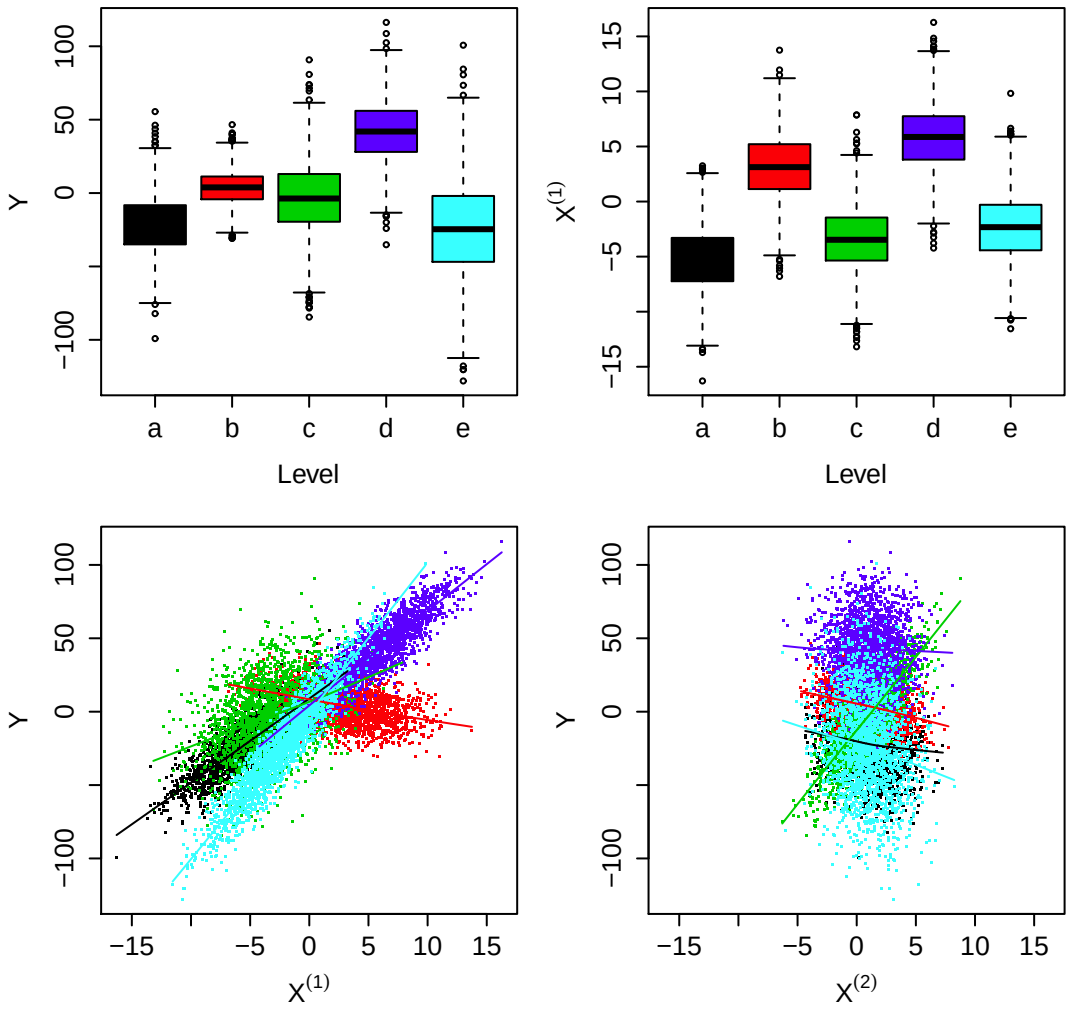

Figure 2.9: Synthetic data: Top panels show box plots of response Y and predictor $X^{(1)}$ by level (a to e) of the random factor (ensemble member). Bottom panels show the relationships between Y and each predictor ($X^{(1)}$ and $X^{(2)}$), coloured by level; curves are local polynomial smoothers.

2.3.2 Impact of dependence

Before leaving this numerical example and discussing the options for specifying a model for the inputs, we examine the impact of dependence between inputs. The inputs have been independent in this numerical example, each from a Gaussian distribution. Now we consider the same statistical model and inputs, except that $X'^{(1)}$ and $X^{(2)}$ are dependent, such that they follow a multivariate Gaussian distribution. We examine the effects on different terms in the decomposition of varying the correlation shared by $X'^{(1)}$ and $X^{(2)}$ between minus one and plus one.

The left panel of Figure 2.10 shows that the relationship between the total variation in Y and the correlation between $X'^{(1)}$ and $X^{(2)}$ is positive and linear. When the correlation is strong and negative the variation in Y is lowest (2389), whereas when the correlation is strong and positive the variation is greatest (up to 2676). This

is because $X'^{(1)}$ and $X^{(2)}$ both (on average) have a positive relationship with Y, and so negative correlation means that the effects of $X'^{(1)}$ and $X^{(2)}$ compensate for each other, whereas positive correlation exacerbates their effect on Y because both inputs tend to be high or low simultaneously.

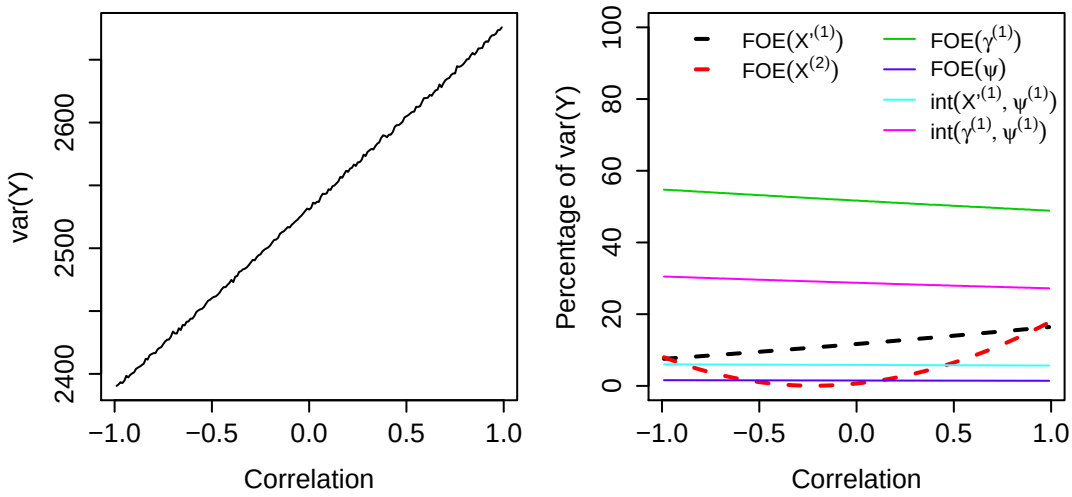

Figure 2.10: Relationship (left panel) between the total variation in Y and the correlation between $X'^{(1)}$ and $X^{(2)}$, and that (right panel) between selected sources of variation and the correlation.

The first-order effects of $X'^{(1)}$ and $X^{(2)}$ are shown by the dashed curves (in black and red respectively) in the right panel of Figure 2.10, expressed as percentages of the total variation in Y. Just as with the total variation in Y, increasing the correlation between $X'^{(1)}$ and $X^{(2)}$ from minus one to plus one has a positive (and linear) effect on the first-order effect of $X'^{(1)}$. This occurs for the same reasons as for the positive relationship between the total variation in Y and the correlation: negative correlation means that the effect of $X^{(2)}$ on Y usually compensates for the effect of $X'^{(1)}$, whereas the opposite occurs when the correlation is positive. By contrast, the first-order effect of $X^{(2)}$ is small when the correlation is weak. The result is that the first-order effect of $X^{(2)}$ is only high when $X^{(2)}$ shares strong correlation with $X'^{(1)}$. When the correlation is close to plus or minus one, the first-order effect of $X^{(2)}$ is similar to that of $X'^{(1)}$ because there is little distinction between them as separate inputs. Therefore the correlation has a different effect on the first-order effect of $X^{(2)}$ because the variation in $X^{(2)}$ is of little importance unless it cannot be disentangled from the variation in $X'^{(1)}$ due to strong correlation.

The positive relationship between the correlation between $X'^{(1)}$ and $X^{(2)}$ and the total variation in Y results in a negative relationship between the correlation and the importance of the other sources of variation shown in the right panel of Figure 2.10. However the interpretations of importance measures (e.g. FOE($\boldsymbol{\psi}$), $\text{var}_{\boldsymbol{\psi}}[E_{\boldsymbol{\mu}}(\boldsymbol{\mu} \mid \boldsymbol{\psi})]$) for which both or neither $X'^{(1)}$ and $X^{(2)}$ are conditioned on are unaffected by corre-

lation between these two inputs because the correlated inputs are grouped together when computing such importance measures (recall from Subsection 1.10.2 and references therein, particularly Jacques et al., 2006).

Overall, Figure 2.10 emphasizes that weak correlation between inputs will not necessarily affect a decomposition to any great extent. In this example, correlation between $X'^{(1)}$ and $X^{(2)}$ weaker than a magnitude of (e.g.) 0.25 has only a small effect on the decomposition relative to the broader conclusions that can be made (e.g. that $\gamma^{(1)}$ is the most important input).

2.4 Model options for inputs

The true model was known in the numerical example of the previous section, and used to demonstrate using simulation to decompose the variation in the response variable. When capturing a driving physical process of precipitation in an ensemble of climate models with a GLMM, however, the true process will be unknown. The parameters of such a GLMM are estimated by fitting it to the empirical data from the climate models, but a model also has to be specified for the inputs. The GLMM does not specify a distribution for all of the inputs, but a distribution is required for each input to compute importance measures. In this section we discuss the options for specifying a model for the inputs, and recommend a general dependence structure.

As discussed in Subsection 2.2.2, when using a GLMM to analyse the representation of a driving physical process of precipitation by climate models, the inputs are 1) the temporal component of each predictor within the climate models, 2) the mean within each climate model of each predictor, and 3) each random effect. Moreover we discussed that the temporal variation in different predictors will often share dependence because the predictors capture the atmospheric conditions represented in the climate models, and that the mean of each predictor and each random effect will often share dependence because they are driven by the differences between the climate models. We therefore proposed that general guidance should be to account for dependence between the temporal variation in different predictors, and between the predictor means and random effects, but that independence can often be assumed between these two sets of inputs. In the rest of this section we discuss the options for specifying a model for the inputs, but we do not restrict the discussion to when this independence assumption is appropriate.

The distributions for the random effects are specified (often as Gaussian or multivariate Gaussian) when specifying the GLMM, and their variance components are

estimated when the GLMM is fitted to the data. Additionally, the value of each random effect can be estimated for each climate model in the ensemble, i.e. their point estimates. Therefore it is natural to use the distributions of the random effects estimated from the GLMM when making inference about the wider population from which the climate models came, and to use the point estimates for the random effects for each individual climate model when making inference only about that specific ensemble. Recall from Section 1.7 that a climate model ensemble can be thought of as one possible collection of models from a wider population of all possible ensemble members, and the interest is often in this wider population.

With respect to the predictor inputs ((1) and (2) above), the distributions can be specified using the empirical data of each input. This is because the mean of each predictor can be removed from the empirical distribution of the predictor in each climate model, resulting in an empirical distribution for the means of each predictor, and an empirical distribution for the deviations from the means for each predictor. If independence can be assumed between a predictor input and the other inputs, then its empirical distribution can be sampled. Alternatively, a kernel density (e.g. see Silverman, 1986, for an overview on kernel density estimates) can be estimated from the empirical distribution of the input. A kernel density estimate is not restricted to the values of the input seen in the empirical data, but does require a bandwidth to be selected (although data-based algorithms are available for automatically selecting the bandwidth, e.g. that of Sheather and Jones, 1991, or see Chacón and Duong, 2010, for a multivariate bandwidth). Another option is to fit a parametric distribution to the empirical distribution of the input, but this requires finding a parametric distribution that fits the data well enough.

If multiple inputs are dependent and it is their combined effect that is of interest (not the individual effect of each), then their joint empirical distribution can be sampled. This is generally not viable, however, if the individual effect of each input is of interest. To see this, suppose that two dependent (continuous) predictors are present ($X^{(1)}$ and $X^{(2)}$) and that we are simulating the first-order effect of $X^{(1)}$. In this case $X^{(1)}$ must be fixed constant, then many values are simulated for $X^{(2)}$ given the value for $X^{(1)}$, before this is repeated many times at different $X^{(1)}$ to generate the required samples. Using the bivariate empirical distribution of the inputs to account for the dependence would result in $X^{(2)}$ being fixed each time $X^{(1)}$ is fixed, and so the individual effect of each could not be disentangled. There are, however, other options for modelling dependence between inputs.

Accounting for linear dependence between inputs can be done by using a multivariate Gaussian copula (e.g. see Nelson, 2006, for an overview on copulae). This

involves transforming each input to be Gaussian (not required if the input is already Gaussian), sampling from a multivariate Gaussian distribution fitted to the transformed inputs, and transforming each input back onto its empirical distribution. This would be a multivariate Gaussian copula with empirical marginals, but instead of transforming back onto the empirical distribution of each of these inputs, they could be transformed onto a kernel density estimate or onto a parametric fit to each empirical distribution. As an illustration, a bivariate Gaussian copula with empirical marginals for two inputs ($X^{(1)}$ and $X^{(2)}$), for when one of the inputs ($X^{(1)}$) is conditioned on, can be implemented by the following procedure:

- Define $Z^{(1)}$ as $\Phi^{-1}\left(\frac{\text{rank}(X^{(1)})}{n+1}\right)$ and $Z^{(2)}$ as $\Phi^{-1}\left(\frac{\text{rank}(X^{(2)})}{n+1}\right)$, where Φ^{-1} is the inverse Gaussian cumulative distribution function, n is the sample size of the empirical data, and $\text{rank}(X^{(1)})$ and $\text{rank}(X^{(2)})$ are the ranks of the empirical vectors of the inputs (i.e. the smallest value of each input is defined as one, second smallest is defined as two, etc). The marginal distributions of $Z^{(1)}$ and $Z^{(2)}$ will be approximately standard Gaussian.

- Randomly sample a value, $Z^{(1)^*}$, for $Z^{(1)}$ from the standard Gaussian distribution.

- Sample the corresponding value, $X^{(1)^*}$, for $X^{(1)}$ by taking the p-th quantile, where $p = \Phi\left(Z^{(1)^*}\right)$, of the empirical distribution of $X^{(1)}$.

- Randomly generate many values for $Z^{(2)}$ from a Gaussian distribution with mean $\rho Z^{(1)^*}$ and variance $1 - \rho^2$, where ρ is the correlation between $Z^{(1)}$ and $Z^{(2)}$.

- Sample the corresponding values, $X^{(2)^*}$, for $X^{(2)}$ by taking the p-th quantile, where $p = \Phi\left(Z^{(2)^*}\right)$, of the empirical distribution of $X^{(2)}$.

A multivariate Gaussian copula accounts for linear dependence between inputs. If dependence is non-linear, however, then non-Gaussian copulae (e.g. see Nelson, 2006) can be considered. These may be attractive at low dimensions, but will often not be flexible enough to account for non-linear dependence between several inputs. Multivariate kernel densities (e.g. see Silverman, 1986) can account for complicated dependence structures between inputs, but can be of limited use at high dimensions because then large spaces can occur between the empirical data-points on which they are based. Vine copulae (see Bedford and Cooke, 2002) are flexible even at high dimensions, although computationally expensive. Another option is to specify a hierarchical dependence structure between the inputs. This could account for complicated dependence structures, although could involve complicated integrals that require Markov chain Monte Carlo methods (e.g. see Brooks et al., 2011) to solve, which can be computationally expensive.

Some methods account for dependence by transforming inputs, using a model to account for the dependence, before transforming the inputs back onto the empirical distribution of each. This is done by the multivariate Gaussian copula with empirical marginals outlined above, for example. Other methods do not require such transformations, such as a multivariate kernel density. There can, however, be advantages to transforming the samples onto the empirical distribution of each input. Firstly, it ensures that values that are physically implausible are not sampled for any individual input. Secondly, transforming onto the empirical distribution of each input ensures that values are sampled within the range of each input on which the GLMM was fitted, and so will generally reduce the extent to which the relationships estimated by the GLMM are extrapolated beyond the data (albeit that samples of multiple inputs can still mean that some such extrapolation is required). Therefore if using a multivariate kernel density to account for dependence, for example, it may advantageous to use empirical marginals by computing the p-th quantile, where

$$p = \frac{\text{rank}(X^*)}{n+1} \, ,$$

of the empirical distribution of input X for each of its samples X^* (n is the number of samples of X^* here). Another option is to constrain the kernel density estimate to be within the empirical range of each input.

As seen above, there are various options when specifying a model for the inputs, and there will often be dependence to account for between some inputs when capturing a driving physical process of precipitation in an ensemble of climate models with a GLMM. However using the general dependence structure proposed in Subsection 2.2.2 (i.e. that dependence between the temporal components of the predictors should be accounted for, as should dependence between the predictor means and the random effects), further general guidance can be offered for specifying a model for the inputs. The dependence between predictor means and random effects may or may not be linear, but accounting for linear dependence between these inputs will at least generally provide a useful first-order approximation. Indeed, detecting any non-linear dependence between these inputs may be difficult anyway because the size of climate model ensembles is often relatively small. As a result, we propose that a multivariate Gaussian copula is generally used for these inputs. Whereas empirical marginals can be used for the variation in each of the means of the predictors, the specified distributions (with the estimated variance components) can be used for the random effects. Alternatively, the point estimates for each ensemble member can be used for the marginal distributions of the random effects. This choice should depend on whether inference is sought about the wider population from which the ensemble members came, or whether the focus is on the specific ensemble members

used.

With respect to the inputs that capture the temporal variation in the predictors, the strength and complexity of the dependence within each climate model between these inputs will depend on how closely they are related in the represented atmospheric conditions, and may also depend on the temporal scale used (e.g. daily-mean data or annual-mean data, etc). As a result, we propose that a multivariate Gaussian copula with empirical marginals is used for these inputs when accounting for linear dependence between them is sufficient. When a more complicated dependence structure is required, we propose that a multivariate kernel density is used with marginals constrained to be within the range of the empirical data. When only a single input is used to capture the temporal variation, its empirical distribution can be sampled. If the assumption that the temporal variation in the predictors is independent from the predictor means and random effects is found not to be valid for a physical process, and that making this assumption does not lead to useful results, then the computationally expensive methods discussed above could be used, such as vine copulae or a hierarchical dependence structure.

Following the guidance suggested above, the empirical data can be used to help determine an appropriate model for the inputs for any given application. For example this can be done by plotting inputs against each other to examine their dependence, or by examining the correlation between them. Moreover, assuming that the size of the empirical data set is large enough, the total variation in precipitation should be estimated to be close to the sample variation, otherwise the implication is that the model for the inputs (or the GLMM) is inappropriate. As such, the estimated total variation in the precipitation can be compared to the sample variation of the empirical data. We do this in the applications of the succeeding chapters.

2.4.1 Impact of modelling decisions: estimating input distributions

Above we discussed the options for using empirical data to specify a model for the inputs. In this, and the next, subsection we highlight that although specifying an appropriate model for the inputs is important for precise results, applications will not necessarily be overly sensitive to the decisions made when modelling the inputs if only approximate results are required. In doing so, we also show that over-fitting should be guarded against when modelling the inputs, particularly the inputs that capture variation across ensemble members when there are few members. To do this we use the numerical example shown in Section 2.3, and the data set generated from the model of this numerical example to visualize the sources of variation in Subsection 2.3.1.

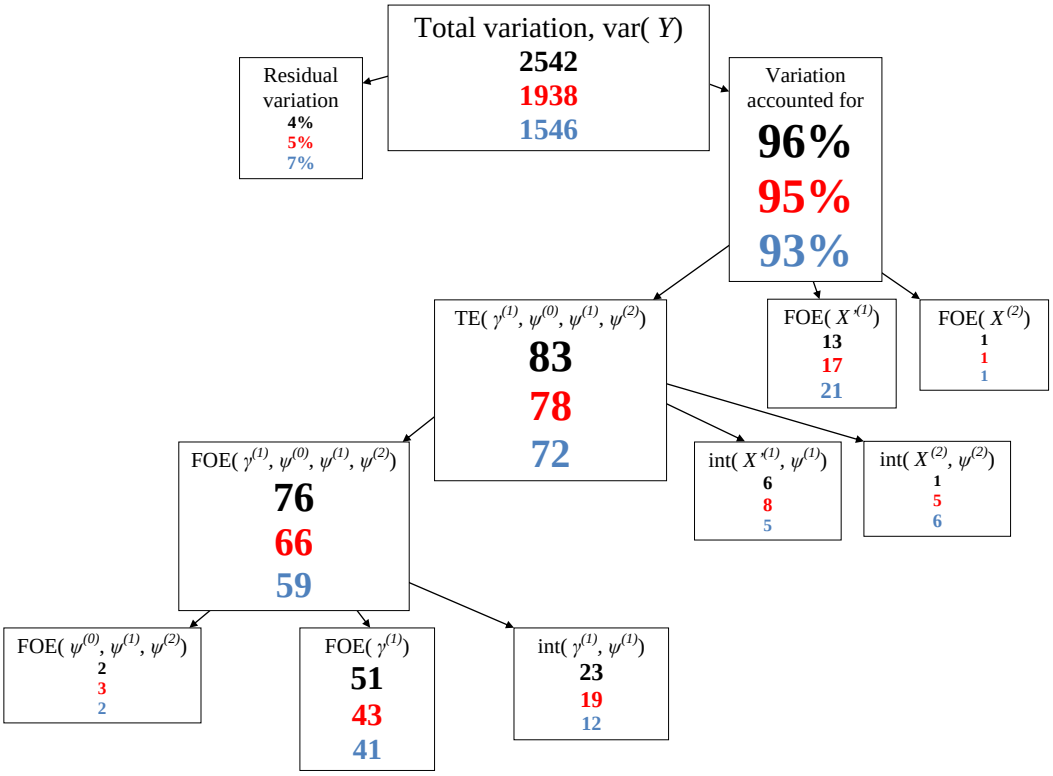

Figure 2.11: Global sensitivity analysis results for the numerical example, showing total effects (TE), first-order effects (FOE) and interactions (int), all expressed as percentages of the total variation in Y. Black values are the true values; red and blue values show simulated estimates based on $R = 1000$ and $r = 2000$ samples from different models for the inputs, see text for details. All values are rounded to zero decimal places.

The black values of Figure 2.11 recall the analytical decomposition of the model used for the numerical example of Section 2.3. If the true model of the inputs was unknown and only the data set generated for Subsection 2.3.1 was available, however, then these data would have to be used to specify the model for the inputs. The red values of Figure 2.11 show the decomposition using simulation when the inputs are modelled using the correct family of distributions (Gaussian in this example) for each input and the correct dependence structure (all inputs independent), but with the non-zero means and variances of the inputs computed from the empirical data, i.e. using plug-in estimates. Showing the true values of these means and variances to the left of the arrows below and the empirical plug-in values to the right of the arrows, these means and variances are

$$\begin{aligned} \psi^{(0)} &\sim (0, 25 \to 26) \\ \psi^{(1)} &\sim (0, 16 \to 18) \\ \psi^{(2)} &\sim (0, 9 \to 28) \end{aligned} \qquad \begin{aligned} \gamma^{(1)} &\sim (1 \to 0.4, 36 \to 22) \\ X'^{(1)} &\sim (0, 9 \to 9) \\ X^{(2)} &\sim (1 \to 1, 4 \to 4) \\ \epsilon &\sim (0, 100 \to 100) \, . \end{aligned}$$

Examining the red values of Figure 2.11, the total variation in response Y is un-

derestimated using the plug-in means and variances shown above. This is largely because the plug-in variance (22) of the mean ($\gamma^{(1)}$) of predictor $X^{(1)}$ is less than that (36) of the true underlying process. The result is that the first-order effect (FOE($\gamma^{(1)}$)) of γ and its interaction (int($\gamma^{(1)}, \psi^{(1)}$)) with its relationship with Y are underestimated. In contrast, the plug-in variance (28) of $\psi^{(2)}$ is greater than that (9) of the true underlying process, resulting in its interaction (int($X^{(2)}, \psi^{(2)}$)) with $X^{(2)}$ being overestimated.

Overall, the plug-in mean and variance estimates for the temporal variation (captured by $X'^{(1)}$ and $X^{(2)}$) are similar to those of the true underlying process because the sample size is relatively large (2000 temporal points for each ensemble member). The discrepancy between the true decomposition and that that uses plug-in estimates is caused by the small number (five) of ensemble members generated for the data set. This is because if the available sample of an input is large enough, then the sample will represent the underlying distribution for that input well, and analysing the data will allow precise inference to be made about the true underlying process. In contrast, however, a small sample size may not represent the underlying process well, in so much as there will be a discrepancy between the results obtained from analysing the data and those that would be obtained if a larger data set was available.

The inputs are known to be independent in the true model, and indeed the correlation between $X'^{(1)}$ and the other inputs, and between $X^{(2)}$ and the other inputs, is 0.0 in the generated data set. Despite this, the small number of ensemble members in the data means that correlation is present between $\gamma^{(1)}$, $\psi^{(0)}$, $\psi^{(1)}$ and $\psi^{(2)}$ in the sample. This correlation is

$$\text{cor}\begin{pmatrix} \gamma^{(1)} \\ \psi^{(0)} \\ \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 0.23 & -0.28 & -0.28 \\ 0.23 & 1 & -0.42 & -0.65 \\ -0.28 & -0.42 & 1 & -0.22 \\ -0.28 & -0.65 & -0.22 & 1 \end{pmatrix}.$$

Now we consider the same decomposition as described above, but with the dependence accounted for using a multivariate Gaussian copula (with the empirical marginal distribution of $\gamma^{(1)}$, and the same Gaussian distributions with plug-in variances described above for the random effects). Using such a Gaussian copula is suggested in the general guidance of Section 2.4 for analysing ensembles of climate models. This decomposition is shown by the blue values in Figure 2.11. In the decomposition, the total variation in Y is underestimated (1546) to a greater extent than when independence was assumed between the inputs (1938). This is because there is negative correlation (e.g. cor($\psi^{(0)}, \psi^{(2)}$) = -0.65 and cor($\gamma^{(1)}, \psi^{(1)}$) = -0.28) between inputs in the data set, and this correlation means that the effects of the

inputs tend to compensate for each other in this model.

There is some discrepancy between the decomposition (blue values) using a multivariate Gaussian copula and that (red values) assuming independence between the inputs. For example the first-order effect (FOE($\gamma^{(1)}, \psi^{(0)}, \psi^{(1)}, \psi^{(2)}$)) of the inputs that capture variation across ensemble members is estimated to be 59% using the Gaussian copula, instead of 66%. This highlights that models for inputs that capture variation across ensemble members can be over-fitted when there are few members.

Despite the discrepancy between the three decompositions shown in Figure 2.11, they are at least qualitatively similar with respect to the broader conclusions that would be drawn. For example $\gamma^{(1)}$ is the most important input in all decompositions, followed by $X'^{(1)}$ and $\psi^{(1)}$. This suggests that applications may not necessarily be overly sensitive to estimating input distributions when only approximate results are required. Indeed, there may even be greater discrepancy between the data set and the true underlying process than the impact of such modelling decisions (i.e. estimation uncertainty can be greater than modelling uncertainty) when the data set is small.

2.4.2 Impact of modelling decisions: estimation methods

With respect to methods of estimating the distributions for the inputs, in this subsection we examine the impact of using the empirical distribution of each input (except for the residual term), compared to unconstrained kernel density estimates, and to kernel density estimates constrained to be within the empirical range of the inputs. As in the previous subsection, we do this by decomposing the variation in the response variable of the data set generated from the model of this numerical example that was used to visualize the sources of variation in Subsection 2.3.1 (recall Figure 2.9, page 71, for a display of the data). We use the empirical distributions of the inputs by sampling with replacement from the marginal distribution of each input to estimate the importance measures, and briefly outline kernel density estimates below.

We use the method of Sheather and Jones (1991) for an automatic bandwidth for each of the kernel density estimates. For a univariate random sample $X_1, ..., X_n$ of size n, a value for bandwidth h is selected for kernel density estimate $\hat{f}_h$, such that

$$\hat{f}_h(x) = n^{-1} \sum_{k=1}^{n} h^{-1} K[h^{-1}(x - X_k)] \, ,$$

where K is the kernel function, which is taken to be a symmetric probability density

(we use the Gaussian family), and x can be any value within the range over which the kernel density is estimated. Therefore a kernel density estimate can be seen as the average of a probability density centred on each data point of the input. Bandwidth h controls the smoothness of the kernel density estimate by controlling the variance of K (i.e. the variance of each probability density that makes up the kernel density estimate). Bandwidth h is chosen to minimize an objective function based on the first two terms of the asymptotic expansion of the mean integrated squared error of the kernel density estimate (see Sheather and Jones, 1991, and references therein for details). Sheather and Jones (1991) use a non-stochastic term to reduce the bias in the estimation without increasing its variance. If a kernel density estimate is unconstrained, then the bandwidth can be used to automatically specify the range of x over which the kernel density is estimated, such that outside this range the kernel density estimate would be negligible anyway (i.e. the estimated probability of X_k being outside this range of x would be nearly zero).

Constrained kernel densities can be estimated by specifying a narrower upper or lower bound for x than would be estimated in the unconstrained case. The kernel density can be constrained like this by setting the probability density for each empirical data point (i.e. each kernel) to zero outside the specified range of x, and correcting for this by inflating the density inside the range of x for that empirical data point by a factor that is the proportion of the density that is outside the range of x. By doing this correction for each kernel separately, instead of for the whole kernel density estimate at once, the density that would be outside of the boundaries in the unconstrained case can remain near (but inside) the boundaries. For example, if a lower bound is set at zero (such that $x \geq 0$) and a data point is observed at zero, then half of the kernel for this data point would be below the lower bound, and so this would be set to zero but the weight of the rest of this kernel would be doubled.

We compare the results obtained from using each of the three estimation methods stated above to those obtained from using a Gaussian distribution with plug-in parameter estimates from the data for each input. The black values of Figure 2.12 show this decomposition, which differs from the decomposition of the red values of Figure 2.11 in the previous subsection because all parameters of the Gaussian distributions are plug-in estimates, including the means of the inputs ($\psi^{(0)}$, $\psi^{(1)}$, $\psi^{(2)}$ and $X'^{(1)}$) for which the true values are known to be zero. This is necessary to make the results comparable with the other estimation methods considered in this subsection. Each estimation method assumes independence between the inputs, which is known to be correct in this example.

The black values of Figure 2.12 show that the total variation (known to be 2542) in

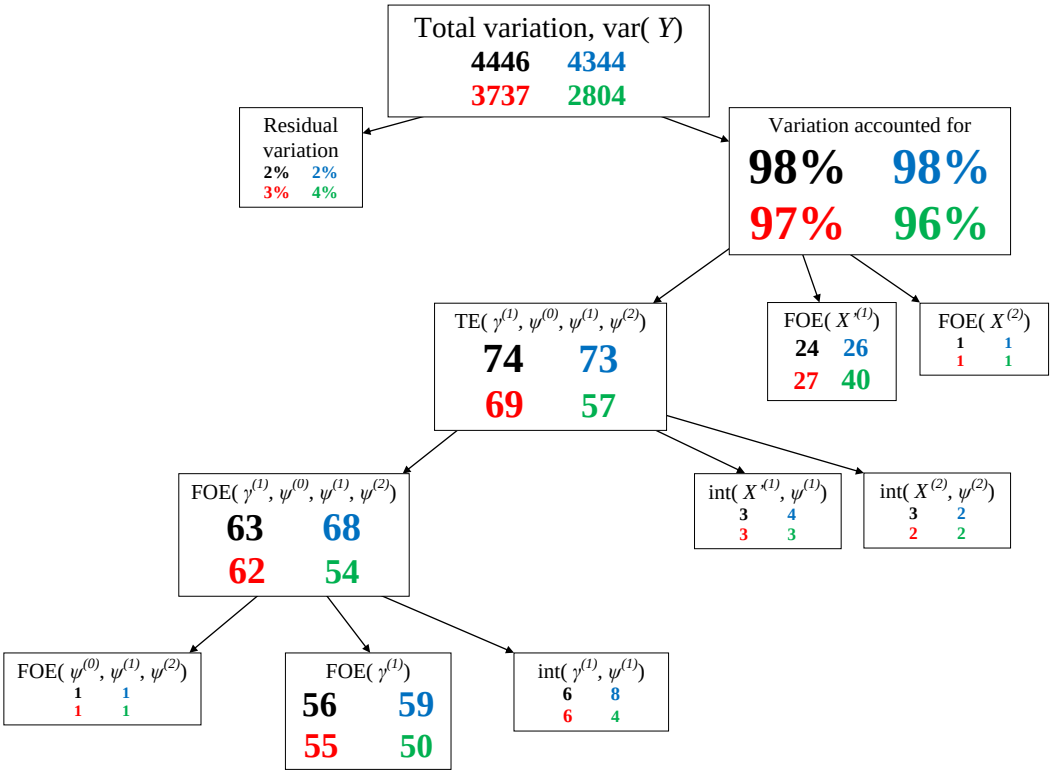

Figure 2.12: Global sensitivity analysis results for the numerical example, showing total effects (TE), first-order effects (FOE) and interactions (int), all expressed as percentages of the total variation in Y. Black values use Gaussian distributions with plug-in parameter estimates to model the inputs, red values use the empirical distribution of each input, blue values use unconstrained kernel density estimates, and green values use constrained kernel density estimates. Each simulation is based on $R = 1000$ and $r = 2000$ samples. All values are rounded to zero decimal places.

response Y is overestimated (4446) when using plug-in estimates for all parameters of the Gaussian distributions. This is largely because the plug-in estimate for the mean of $\psi^{(1)}$ is 4.8, i.e. the mean relationship between $X^{(1)}$ and Y is estimated to be greater than the true underlying process (recall that $E(\psi^{(1)}) = 0$ in the known true process), due to the small number (five) of generated ensemble members. With respect to the rest of the decomposition, the main result of this is that the variation solely due to that in $X^{(1)}$ is overestimated, but the differences between ensemble members are underestimated (due to the underestimation of $\text{int}(X^{(1)}, \psi^{(1)})$) relative to the estimated total variation in Y.

The other estimation methods considered in this subsection can be compared to the black values of Figure 2.12 because this decomposition uses Gaussian distributions, which are true to the known underlying model in this example, and using plug-in estimates is appropriate here because we are applying different estimation methods to the empirical data set. Compared to the black values of Figure 2.12, the red values show that restricting the values taken for each input by sampling from their (marginal) empirical distributions results in the total variation in Y being underestimated (3737 compared to 4446) in this example. Despite this, the rest of the

decomposition is similar (albeit that TE($\gamma^{(1)}, \psi^{(0)}, \psi^{(1)}, \psi^{(2)}$) is underestimated and FOE($X'^{(1)}$) is slightly overestimated).

Unlike when sampling from the empirical distribution of each input, using an unconstrained kernel density estimate (blue values of Figure 2.12) for each input estimates the total variation in Y to be similar to when Gaussian distributions are used (4344 compared to 4446). The rest of the decomposition is also similar (the largest discrepancy is 5 percentage points for the estimate of FOE($\gamma^{(1)}, \psi^{(0)}, \psi^{(1)}, \psi^{(2)}$)). The decomposition shown by the green values of Figure 2.12 uses a kernel density estimate for each input constrained to be within the empirical range of the input, and is substantially different (e.g. FOE($X'^{(1)}$) = 40) to those obtained using the other estimation methods. This is because constrained kernel densities underestimate the variation in each input, which substantially underestimates the importance of the inputs that capture variation across different ensemble members because these inputs only have five empirical values each.

Overall, it can be seen from Figure 2.12 that constrained kernel density estimates should not be used for inputs with few available empirical samples, but that using unconstrained kernel density estimates and empirical distributions can lead to similar decompositions, even when few ensemble members are present. This subsection supports the general recommendation provided of Section 2.4, which suggests using the fitted distributions from the GLMM for modelling the random effect inputs (not constrained kernel density estimates for these inputs). Moreover, recalling the black values of Figure 2.11 in Subsection 2.4.1, the impact of using different estimation methods can be less than the discrepancy between the data set and the true underlying process, i.e. less than the estimation uncertainty. Along with the previous subsection, this shows that applications may not necessarily be overly sensitive to decisions made for modelling inputs, particularly when only approximate results are required.

2.5 Random-effects ANOVA revisited

Now that the methodology has been outlined, it is worth relating the general decomposition of Subsection 2.2.1 to the random-effects analysis of variance (random-effects ANOVA) described in Section 1.7. In particular, it is worth examining the advantages of using importance measures with a GLMM that captures the representation of a driving physical process of precipitation in an ensemble of climate models.

Recall from Section 1.7 that a random-effects ANOVA estimates the variation across

ensemble members in the mean rate of precipitation. It does this because no predictor variables are included, and the only input (other than the residual term) is a single random effect (ψ^*) that captures the variation across ensemble members in the expectation μ of the precipitation Y, i.e.

$$\begin{aligned} Y_{ij} \mid \psi_j^* &\sim (\mu_j, \tau_\epsilon) \\ \mu_j &= \beta_0 + \psi_j^* \\ \psi_j^* &\sim (0, \tau_0) \ , \end{aligned}$$

where i indexes time, j indexes ensemble member, β_0 is a fixed parameter, and τ_0 and τ_ϵ are the variance components of ψ^* and the residual term respectively.

Since the random-effects ANOVA uses τ_0 to estimate the variation across ensemble members in the expectation of precipitation, we can express this in terms of importance measures, and relate it to the sources of variation captured by a GLMM. The variation across ensemble members in mean precipitation estimated by a random-effects ANOVA can be expressed as $\text{var}_{\psi^*}[E_Y(Y \mid \psi^*)]$, i.e. the first-order effect of ψ^*. This is equivalent to $\text{var}_{\psi^*}[E_\mu(\mu \mid \psi^*)]$, by the definition of μ. This is also the total variation found to be due to the differences between ensemble members when using a random-effects ANOVA. This is clear because ψ^* is the only input other than the residual term in the random-effects ANOVA, and so ψ^* is the only input that captures variation across the different members, and there are no inputs with which ψ^* could interact (so its first-order effect equals its total effect).

In capturing variation across ensemble members in mean precipitation, the random effect ψ^* captures the variation in the intercept of the ANOVA. However as discussed previously in Section 2.1, the physical interpretation of the intercept of a statistical model depends on the predictors that are included. Therefore, the intercept of a random-effects ANOVA will generally be different to the intercept of a GLMM, and so FOE(ψ^*) will generally be different to the first-order effect of the variation in the intercept of a GLMM.

Despite the difference between FOE(ψ^*) and the first-order effect of the intercept of a GLMM, FOE(ψ^*) can be recovered by using importance measures and a GLMM. This is because FOE(ψ^*) can be recovered by the first-order effect (FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$)) of the predictor means ($\boldsymbol{\gamma}$) and random effects ($\boldsymbol{\psi}$) in the GLMM, i.e. the first-order effect of the inputs to a GLMM that capture the variation across the different ensemble members. To see this, consider the mathematical definition of FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$), $\text{var}_{\boldsymbol{\gamma},\boldsymbol{\psi}}[E_\mu(\mu \mid \boldsymbol{\gamma}, \boldsymbol{\psi})]$. This is the variation across ensemble members in mean precipitation rate remaining after averaging over the temporal variation $\boldsymbol{X'}$ in the

predictors. There are no predictors present in a random-effects ANOVA to capture this temporal variation, and so it is automatically averaged over when examining the effect of ψ^*. As such, the definition of FOE(ψ^*) matches that of FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$).

Despite having equivalent definitions, FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) and FOE($\psi^*$) may not be estimated to be precisely equal, of course. Even though FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) and FOE($\psi^*$) are estimates for the same concept, they use different statistical models and so different estimators. As a result, the appropriateness of the assumptions that underpin the statistical models will affect the precision of the estimates (e.g. a poorly fitting statistical model will not result in precise estimates).

Recognising that FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) is equivalent to the result from a random-effects ANOVA highlights the additional information that can be gained from capturing the representation of a driving physical process of precipitation using a GLMM. Firstly, comparing FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) to the total effect (TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$)) of the same inputs reveals the additional variation in the simulation of precipitation that is found to be due to the differences between ensemble members. This additional variation is due to the interactions between the inputs ($\boldsymbol{\gamma}$ and $\boldsymbol{\psi}$) that capture variation across ensemble members and those ($\boldsymbol{X'}$) that capture temporal variation, i.e. int($\boldsymbol{X'}, \boldsymbol{\gamma}$), int($\boldsymbol{X'}, \boldsymbol{\psi}$) and int($\boldsymbol{X'}, \boldsymbol{\gamma}, \boldsymbol{\psi}$).

The second advantage of using a GLMM over a random-effects ANOVA is that insight can be gained into what causes the variation across and within the ensemble members. For example it could be that the variation found to be due to the differences between ensemble members is dominated by the variation in the frequency distribution of a single predictor, or by the variation in the relationship between a predictor and the precipitation. This could be identified by using a GLMM to decompose the variation in the representation of a driving physical process of precipitation. In contrast, the causes of the variation quantified by a random-effects ANOVA are undiagnosed because the variation is only captured by a single input (ψ^* above).

2.6 Limitations

In Subsection 2.2.2 it was seen that dependence between the random effects and the predictor means can make some of the decomposition non-orthogonal. Although weak dependence will not substantially affect the decomposition, strong dependence could make this non-orthogonality so great that the effects of different inputs cannot be disentangled at all. This would limit the detail gained from the decomposition. Some analyses will not require such detail, and indeed revealing the dependence

structure may be useful in itself. However strong dependence between more of the inputs, including between the temporal variation in the predictors and the other inputs, could severely limit the use of the methodology. Dependence between temporal variation in predictors and inputs that allow variation across ensemble members could mean that there is variation across the members in the shape of the distributions of the predictors (not just their means). This would result in more of the decomposition being complicated by dependence between inputs because the effect of the temporal variation could not be fully disentangled from that of the variation due to differences between the ensemble members.

In addition to strong dependence between many inputs limiting the use of the methodology, accounting for such dependence could require specifying a complicated dependence structure. For example if there is variation across ensemble members in the shape of the distribution of a predictor, then accounting for this dependence could require a complicated model for the inputs. Simulating from a complicated dependence structure could make the methodology computationally expensive. This may not be a problem when running the simulation once, but applications using climate models often require analysing data at multiple grid points. Running a computationally expensive simulation at many grid points may not be feasible. When strong dependence or a complicated dependence structure limits the use of the methodology, different predictors may have to be sought to capture the representation of the physical process. For example it may be that two strongly dependent predictors could be replaced by three weakly dependent predictors that capture similar atmospheric conditions.

The methodology requires using a GLMM with inputs sampled from a specified model. Therefore if inputs are sampled from outside the (multivariate) range on which the GLMM was fitted, then the GLMM must be used to extrapolate outside of this range. In this case judgement must be used to determine whether or not such samples lead to a realistic sample for the response variable. GLMMs can be particularly sensitive when extrapolating in the presence of a non-linear link function (e.g. the log link). A way of reducing such sensitivity could be to specify a different link function for any extrapolation outside the range of the fitted values of the GLMM. For example by finding a way to smoothly (continuously) extrapolate a log function into a linear function outside the range of the fitted values (this is done, and detailed, in Chapter 4). Unrealistic sampling means that the model for the inputs, or perhaps the specification of the GLMM, must be altered.

2.7 Applications

In the following chapters the methodology proposed in this chapter is applied to two different climate model ensembles, an ensemble of general circulation models (GCMs) and a perturbed physics ensemble of a regional climate model (RCM). In Chapter 3 the methodology is used to investigate change in temperature in response to increased CO_2 and its influence on the change in global-mean precipitation in the ensemble of GCMs, and this is compared to a similar analysis over the tropical ocean. As discussed in the next chapter, the statistical modelling required in Chapter 3 is relatively simple, the advantages of using our methodology over simpler methods is limited in the application, and the results complement those of Lambert and Webb (2008), rather than adding to them. Therefore the application is not designed to add much to scientific knowledge, but instead we consider the analysis to be a preliminary analysis that serves to demonstrate the methodology using real data from climate models, and on a research topic in which the results can be compared to those of Lambert and Webb (2008).

In Chapter 4 we analyse the extent to which the simulation of daily-total summer convective precipitation is robust to uncertainty in the parameters of an RCM at the grid-point level over southern Britain. In doing so, we also show that the methodology can be used with a complicated GLMM and to produce maps to examine the spatial pattern of importance measures. We then demonstrate how the influence of perturbing individual RCM parameters can be quantified in Chapter 5. In Chapter 6 we extend the methodology to focus on extreme precipitation, and use the extended methodology to analyse the extent to which extreme large-scale precipitation events are simulated robustly. Chapter 7 summarizes the thesis, and provides final discussion and ideas for future work.

3 Global and tropical-ocean precipitation

Aim: To quantify the extent to which, and understand why, the simulated change in global-mean and tropical ocean-mean precipitation in response to increased CO_2 varies across GCMs.

3.1 Introduction

Temperature plays an important role in the driving physical processes of precipitation, with warmer air able to hold more moisture (e.g. see Allen and Ingram, 2002) and a greater vertical temperature gradient (lapse rate) through the atmosphere leading to greater convection. Lambert and Webb (2008) used general circulation models (GCMs) to analyse the causes behind the effect of surface temperature on global-mean annual-mean precipitation after a doubling of carbon dioxide (CO_2), and found that these causes are dominated by processes robustly simulated by GCMs. In particular, they found that GCMs simulate the relationship between change in temperature and change in global-mean precipitation more robustly than the change in global-mean temperature in response to increased CO_2. This is because change in global-mean precipitation is constrained by the tropospheric energy budget (Allen and Ingram, 2002). The dominant processes of the tropospheric energy budget are clear-sky emission, absorption of longwave radiation, and absorption of shortwave radiation, and these processes are relatively well-understood compared to shortwave cloud feedbacks (see Ramanathan et al., 2001, and Lambert et al., 2008, as well as Lambert and Webb, 2008).

With respect to the language that we use throughout this chapter, we simplify 'change in precipitation in response to increased CO_2' to 'precipitation change' (and similar for 'temperature change'). In the chapter we complement the work of Lambert and Webb (2008) by using the methodology of Chapter 2 to examine the extent to which the simulation of global-mean precipitation change in response to increased CO_2, including the dependence of the precipitation on temperature, is robust across GCMs. In doing so, we quantify the proportion of the total variation in global-mean annual-mean precipitation change in an ensemble of GCMs that is due to the differences between the GCMs. We also compare the contributions to this of the variation across GCMs in the relationship between temperature change and precipitation change, and of the variation across the GCMs in the distribution of temperature change. It is because Lambert and Webb (2008) have already investigated these sources of variation that the purpose of this chapter is to demonstrate the methodology of Chapter 2 using real climate model data and complement the work

of Lambert and Webb (2008), rather than to add much to the scientific knowledge on the topic.

There is likely to be greater agreement between GCMs in their simulation of global-mean precipitation than when the precipitation is averaged over a smaller region because there are greater physical constraints on total global precipitation. Global energy conservation constrains the equilibrium global-temperature change after a given forcing, the Clausius-Clapeyron relation determines the effect of this on the capacity of the air to hold moisture, and the tropospheric energy budget constrains the resulting changes in global precipitation (Mitchell et al., 1987; Allen and Ingram, 2002; also see Lambert and Webb, 2008). Therefore it may be of interest to find the region over which there is least agreement between GCMs in their simulation of precipitation, and to use it with the global analysis to examine the extent to which the agreement depends on spatial region.

Figure 3.1 reproduces Figure SPM.8(b) from Working Group I of the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (2013). The figure shows projected patterns of change in annual-mean precipitation across the globe using two different future greenhouse gas emission scenarios. In the figure, hatching indicates regions where the multi-model mean is small compared to natural internal variability, whereas stippling indicates regions where the multi-model mean is large compared to natural internal variability and where at least 90% of models agree on the sign of change. It can be seen that there is more agreement between climate models at high latitudes than in the tropics in both emission scenarios shown (despite some stippling in the tropics in the right panel, there is much hatching in and around the tropics, unlike at high latitudes). Furthermore there is generally less agreement between climate models in their projections of precipitation change at mid-latitudes in the summer than in the winter in each hemisphere (see Figures 14.18 to 14.24 from Chapter 14 of Working Group I of the IPCC Fifth Assessment Report by Christensen et al., 2013; hatching implies model agreement in these figures). Convective precipitation is most prominent in the tropics, and is more prominent in the summer than winter at mid-latitudes. Convection is difficult to represent within GCMs because of the fine resolutions on which it occurs, which results in the GCMs representing convection by parameterizations and different convection schemes rather than being resolved explicitly (see Vidale et al., 2003, Déqué et al., 2007, and Hohenegger et al., 2008).

The difficulty in representing convection contributes to the disagreement between GCMs when convective precipitation is prominent, as well as continental-scale bias in the tropics (Marsham et al., 2013; Birch et al., 2014a; also see Stephens et al.,

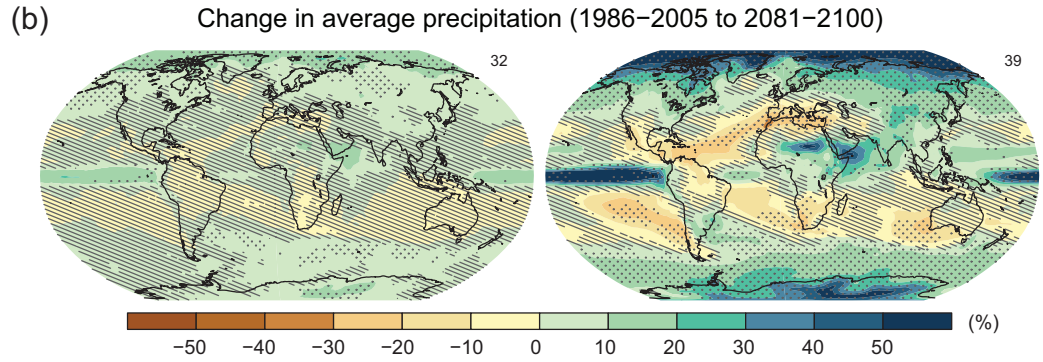

Figure 3.1: Figure SPM.8(b) from Working Group I of the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (2013). Left panel uses the RCP2.6 emissions scenario (see Moss et al., 2010); right panel uses RCP8.5. The number of climate models used to calculate the multi-model mean is indicated in the upper right corner of each panel. Hatching indicates regions where the multi-model mean is small compared to natural internal variability (less than one standard deviation of natural internal variability in 20-year means). Stippling indicates regions where the multi-model mean is large compared to natural internal variability (greater than two standard deviations of natural internal variability in 20-year means) and where at least 90% of models agree on the sign of change.

2010, and Birch et al., 2014b). This, in addition to the greater physical constraints on global precipitation, suggests that the total variation in precipitation change that is due to the differences between the GCMs is greater over the tropics than globally. Moreover, this signal is likely to be stronger when comparing global precipitation to the tropical *ocean* precipitation. This is because only considering ocean grid-points over the tropics removes any limitation from moisture availability and any orographic effects, thereby allowing the impact on precipitation of the variation across GCMs in the representation of convection to dominate to a greater extent. Therefore the disagreement between GCMs in their simulation of precipitation is likely to be greater over the tropical ocean than other regions. As such, we can compare our analysis of global-mean annual-mean precipitation to a similar analysis of annual-mean precipitation averaged over the tropical ocean, and thereby estimate something close to an upper bound for the disagreement between GCMs in their simulation of the change in spatially-averaged annual-mean precipitation in response to increased CO_2. This cannot be seen as a formal upper bound, but the theory outlined above suggests that it may be difficult to find a region where there is greater disagreement between GCMs.

3.2 Data

We follow Lambert and Webb (2008) in using GCM experiments in which CO_2 is increased abruptly, and temperature and precipitation are anomalized with respect to a control period for each GCM before the CO_2 forcing (i.e. temperature change and precipitation change are taken relative to a control period in each GCM, we

detail the data processing below). Whereas Lambert and Webb (2008) used experiments in which CO_2 is doubled, we use experiments in which the CO_2 concentration is instantaneously quadrupled. The strength of the response of temperature and precipitation to the forcing should be stronger in these experiments. Specifically, the data are from the abrupt 4xCO_2 (150 years) experiments for the fifth coupled model intercomparison project (CMIP5) (see Taylor et al., 2012). The control period in each GCM for these experiments is for at least 250 years of pre-industrial conditions, and for more than 500 years in most of the GCMs. We use an ensemble of fourteen GCMs for which there are 150 years of data after the CO_2 forcing for each GCM. We index the GCMs by j, and list them in Table 3.1.

Index (j)	GCM	Modelling centre/group
1	bcc-csm1.1	Beijing Climate Center, China Meteorological Administration
2	CanESM2	Canadian Centre for Climate Modelling and Analysis
3	CCSM4	National Center for Atmospheric Research
4	CNRM-CM5	Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique
5	CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence
6	GFDL-CM3	NOAA Geophysical Fluid Dynamics Laboratory
7	GISS-E2-H	NASA Goddard Institute for Space Studies
8	HadGEM2-ES	Met Office Hadley Centre / Instituto Nacional de Pesquisas Espaciais
9	IPSL-CM5A-LR	Institut Pierre-Simon Laplace
10	MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology
11	MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies
12	MPI-ESM-LR	Max-Planck-Institut für Meteorologie (Max Planck Institute for Meteorology)
13	MRI-CGCM3	Meteorological Research Institute
14	NorESM1-M	Norwegian Climate Centre

Table 3.1: List of GCMs.

The temperature that we use is near-surface air temperature. To anomalize this with respect to the control period in each GCM, we initially anomalize by month at each grid point separately and each GCM separately. We do this by taking the difference between each monthly-mean temperature in January after the CO_2 forcing and the mean temperature in all Januarys before the forcing, for each grid point for each GCM. This is also done separately for the remaining months, February to December. We then aggregate up to change in annual-mean temperature at each

grid point for each GCM. After this we average spatially (weighting the grid points by their surface area) for each GCM. For the tropical ocean data, we only use grid points over the ocean between $+/-30$ degrees latitude. This gives us the change in annual-mean global-mean temperature in response to the CO_2 forcing, and the change in annual-mean tropical ocean-mean temperature in response to the forcing. The same processing is done for precipitation.

3.3 Exploratory data analysis

Figure 3.2 displays the time series of the data, separated by GCM. The spatially-averaged temperature change (ΔT) and precipitation change (ΔY) are measured in Kelvin (K) and millimetres (mm) per day respectively. The time series all show an initial upward trend followed by a levelling off as the climate converges towards its equilibrium in each GCM. Contrasting the global and tropical-ocean temperature, the global temperature tends to warm more than over the tropical ocean. The variation across GCMs in the distribution of the temperature changes appears similar globally and over the tropical ocean. With respect to precipitation, the variability is greater over the tropical ocean than globally both across the GCMs and internally within each GCM. The mean precipitation change across the ensemble members appears similar in both data sets. Whereas the effect of quadrupling CO_2 on temperature is an immediate increase, precipitation increases but only after an initial decrease. This initial decrease in precipitation after the forcing is of greater magnitude over the tropical ocean than globally.

Whereas we could focus on (something close to) the equilibrium climate in each GCM by (e.g.) only using the last 50 years of each time series, we choose to use the whole 150 years of each time series. We do this because, as shown below, including the spin up period (i.e. the strong upwards trend in the early years) in each GCM does not complicate the relationship between the change in temperature and precipitation (the relationship is approximately linear after the spin up period, as well as during it), and by including the spin up periods we are able to analyse the entire response of the precipitation to the forcing, including examining the variation across GCMs in the initial decrease in precipitation seen in Figure 3.2 immediately after the forcing. As we discuss later in the chapter (in its final section), the differences between the GCMs are more prominent towards the end of the time series than near the start. This can be seen in Figure 3.2 because the ensemble spread widens as time progresses after the forcing, and means that a result of including the spin up periods is that the differences between the GCMs are less prominent than they would appear if only the latter part of the time series were used. However, as seen later by the results of the chapter, this does not prevent the purposes of the chapter from being achieved,

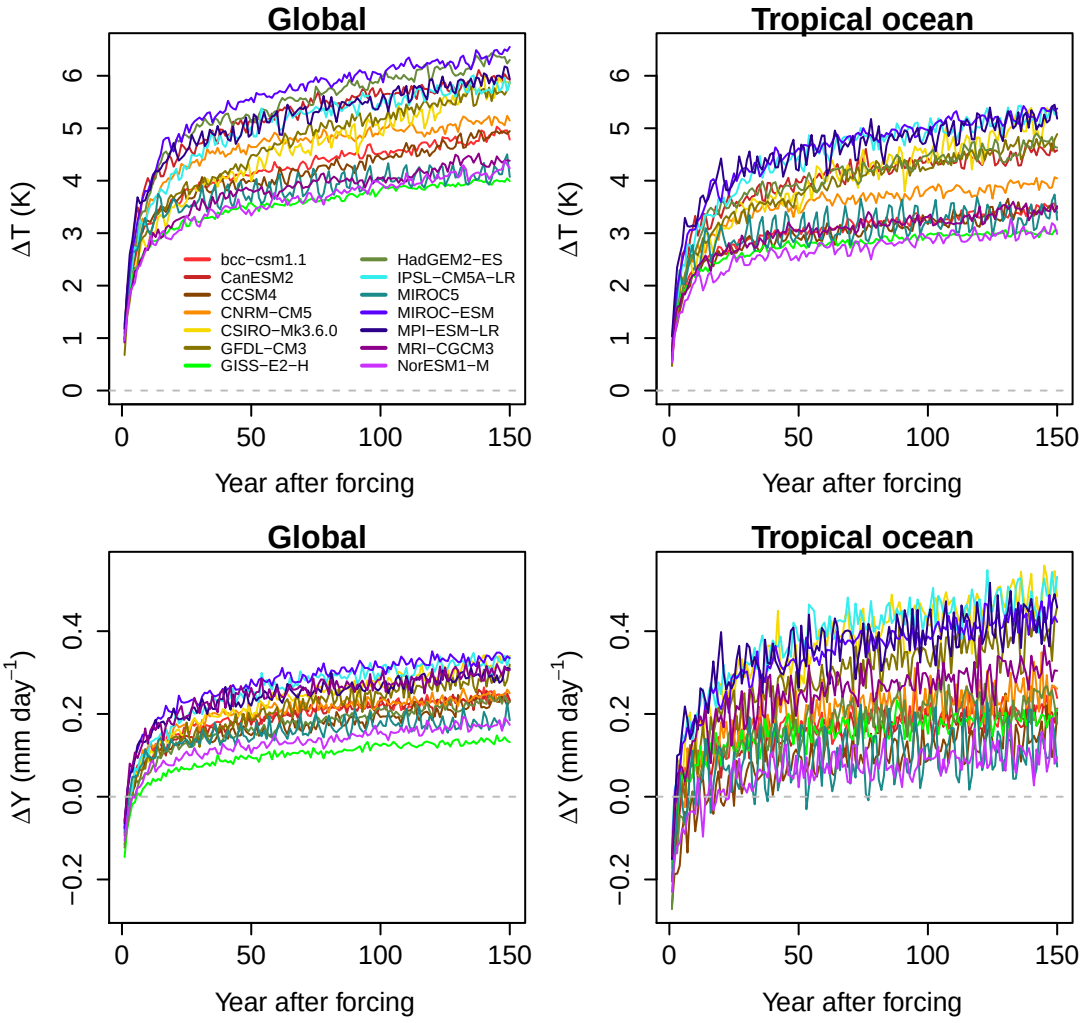

Figure 3.2: Change (from control period) in global-mean (left panels) and tropical ocean-mean (right panels) annual-mean temperature (ΔT) (top panels) and precipitation (ΔY) (bottom panels) times series, by GCM.

which are to demonstrate the methodology of Chapter 2 using real climate model data and to complement the work of Lambert and Webb (2008).

The initial decrease in precipitation after the forcing is due to ocean heat uptake (McInerney and Moyer, 2012). After an abrupt CO_2 forcing there is an instantaneous positive net radiative imbalance at the top of the atmosphere, resulting in a downward energy flow through the atmosphere and ocean. However the greater thermal inertia of the ocean means that the ocean heats more slowly than the atmosphere, thereby suppressing warming at the surface of the ocean and creating a differential between the warming at the surface and the warming higher in the atmosphere. This increases the convective stability of the atmosphere, reducing convection and therefore precipitation. See McInerney and Moyer (2012) and references therein for further detail. Since the initial decrease in precipitation after the forcing is due to ocean heat uptake, it is not surprising that it is of greater magnitude over the

tropical ocean than globally because ocean heat uptake affects convection (which is more prominent over the tropics) and occurs over the ocean.

No GCM can be seen as remarkably different from the others from Figure 3.2, although (for example) MIROC-ESM tends to simulate high temperature and precipitation, and GISS-E2-H tends to simulate low temperature and precipitation. Indeed, the GCMs appear to be consistent over the two different spatial scales, with GCMs simulating high (low) anomalies globally also simulating high (low) anomalies over the tropical ocean. Furthermore the GCMs with high (low) temperature anomalies tend to have high (low) precipitation anomalies.

A potential hazard of relating variables with time trends to each other is that if changes over time in two variables are driven by a different process then correlation can be induced between the variables when in reality they do not have a causal link, i.e. it is more likely that correlation is confused for causation. However in this case the causality between temperature and precipitation has a physical basis. The capacity of air to hold moisture depends on temperature, and temperature drives convection, see Section 3.1. Moreover, we are using the $4xCO_2$ experiments because it is known that temperature will increase after quadrupling the concentration of carbon dioxide in the atmosphere, and it is largely this temperature increase that drives the changes in precipitation.

Figure 3.3 shows precipitation change against temperature change globally and over the tropical ocean. The points are tight to the curves, suggesting that temperature accounts for much of variation in the precipitation. There appears some curvature in the relationship in some GCMs, but assuming linearity in each GCM would approximate the relationship between the temperature and precipitation well. There is perhaps more variation across GCMs in the relationship over the tropical ocean than globally. The points seem evenly and symmetrically scattered around each curve without any obvious changes of variance. No outlying values appear present in the data.

From Figure 3.3, there appears little variation across the GCMs in the relationship between temperature change and precipitation change globally, i.e. the slopes appear similar. Despite this, (for example) MRI-CGCM3 tends to simulate high precipitation changes given temperature, and a strong relationship between temperature and precipitation contributes to this. Similarly, HadGEM2-ES tends to simulate low precipitation changes given temperature, and a weak relationship between temperature and precipitation contributes to this. A high intercept (i.e. the initial decrease in precipitation immediately after the forcing) also contributes to the high precipitation

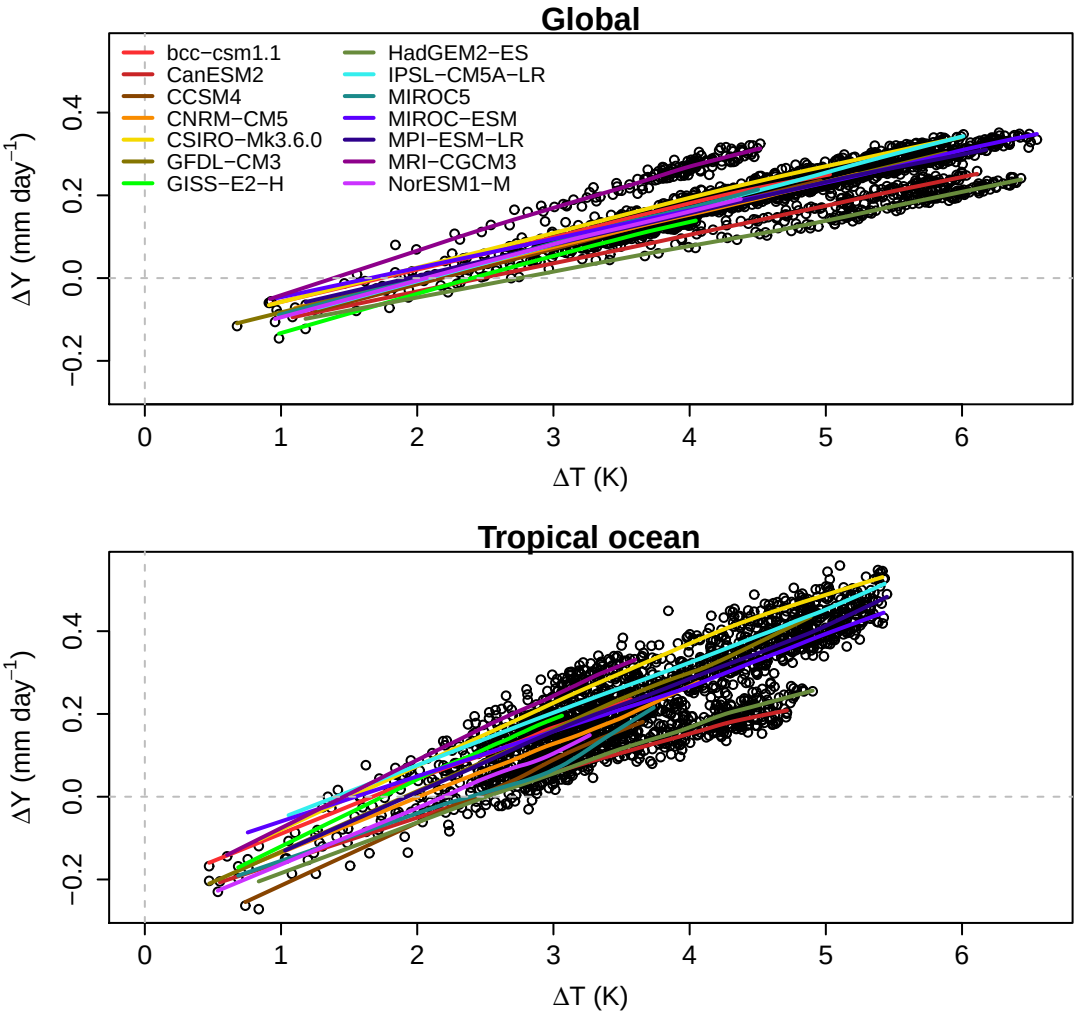

Figure 3.3: Change in global-mean (top panel) and tropical ocean-mean (bottom panel) annual-mean precipitation (ΔY) against change in temperature (ΔT). Curves are local polynomial smoothers for each GCM.

changes given temperature in MRI-CGCM3, and a low intercept contributes to the low precipitation changes given temperature in HadGEM2-ES. However it appears that low temperature changes are simulated by MRI-CGCM3, thereby compensating for its strong relationship and high intercept, resulting in the model simulating lower precipitation changes than it otherwise would. Similarly, it appears that high temperature changes are simulated by HadGEM2-ES, compensating for its weak relationship and low intercept. As a result, the precipitation changes simulated by MRI-CGCM3 are not greater than those simulated by many of the other GCMs, and the precipitation changes simulated by HadGEM2-ES are not less than those simulated by many GCMs (also see bottom-left panel of Figure 3.2).

The variation across GCMs in the intercept seen in the top panel of Figure 3.3 indicates that there is variation across GCMs in the initial decrease (due to ocean heat uptake) in precipitation after the CO_2 forcing. Over the tropical ocean, there

appears to be more variation across GCMs in the relationship between temperature change and precipitation change (see bottom panel of Figure 3.3). However this is perhaps only because of the relatively weak relationships in CanESM2 and HadGEM2-ES. There also appears more variation across GCMs in the initial decrease in precipitation after the forcing over the tropical ocean than globally.

Figure 3.4 shows box plots of the changes in temperature and in precipitation. They confirm that there is considerable variation across the GCMs in the distributions of the changes in temperature and precipitation both globally and over the tropical ocean. The box plots highlight that there is greater variation across the GCMs in the distribution of the precipitation changes over the tropical ocean than globally, but less difference between the two data sets in the variation across the GCMs in the distribution of the temperature changes. The greatest difference between the two data sets in the distributions of the temperature changes is that they are higher globally than over the tropical ocean. The distributions of both the temperature changes and those in precipitation are negatively skewed due to the shape of their time series: the strong increase in both variables over the years immediately after the CO_2 forcing, followed by a levelling off as they converge towards their post-forcing equilibrium distributions.

Overall, the exploratory data analysis has shown that much of the variation in the changes in precipitation in response to increased CO_2 can be accounted for by the variation in the changes in temperature. As well as the temporal variation in temperature change, there is considerable variation across GCMs in the distribution of the temperature changes. Some of the variation in precipitation change seems to be due to variation across the GCMs in the initial decrease in precipitation after the forcing (due to ocean heat uptake, see above), particularly over the tropical ocean. There seems little variation across the GCMs in the relationship between temperature change and precipitation change, particularly globally. To assess these sources of variation, we follow the methodology of Chapter 2 by fitting a statistical model to the data.

3.4 Statistical modelling

In light of the exploratory data analysis above, the specification of the statistical model used for each of the two data sets is

$$\Delta Y_{ij} \mid \left(\psi_j^{(0)}, \psi_j^{(1)}, \Delta T_{ij}\right) \sim \text{Normal}(\mu_{ij}, \sigma^2)$$

$$\mu_{ij} = \beta_0 + \beta_1 \Delta T_{ij} + \psi_j^{(0)} + \psi_j^{(1)} \Delta T_{ij}$$

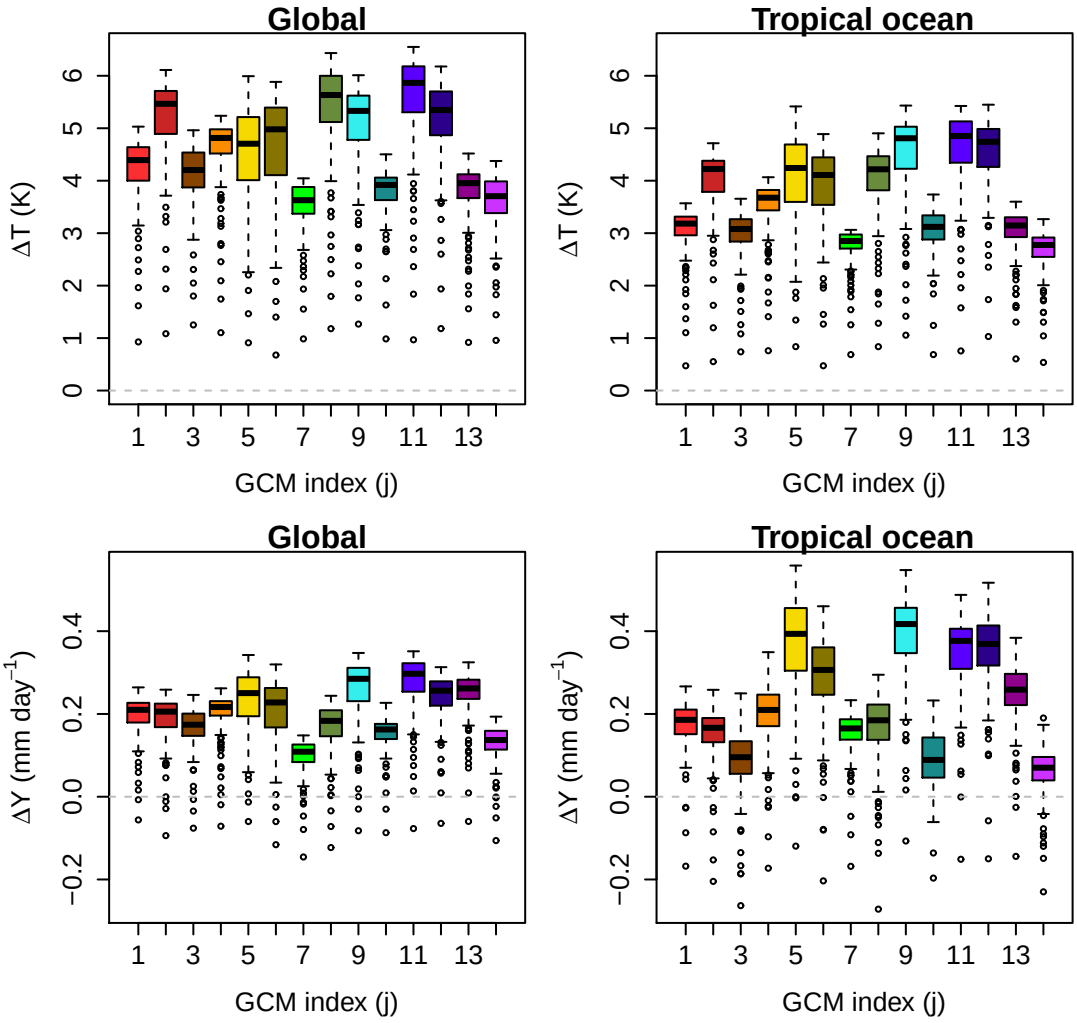

Figure 3.4: Box plots of change in global-mean (left panels) and tropical ocean-mean (right panels) annual-mean temperature (ΔT) (top panels) and precipitation (ΔY) (bottom panels) by GCM.

$$\boldsymbol{\psi} = \left(\psi_j^{(0)}, \psi_j^{(1)}\right) \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}) \ .$$

As the specification shows, we relate the spatially-averaged change in annual-mean near-surface air temperature (ΔT, in Kelvin) linearly to the spatially-averaged change in annual-mean precipitation (ΔY, in mm per day) in response to increased CO_2 using a general linear mixed model (LMM) (see Breslow and Clayton, 1993). Two multivariate Gaussian random effect terms ($\psi^{(0)}$ and $\psi^{(1)}$) are used to allow the intercept and the slope to vary with GCM, and the residual term is assumed to be Gaussian with constant variance. The fitted expectation of precipitation change is denoted by μ, i indexes the year within each GCM (and recall that j indexes the GCM), and σ^2 and $\boldsymbol{\Sigma}$ are the variance components of the residual term and the random effects respectively.

We use restricted maximum likelihood (REML) (Patterson and Thompson, 1971)

to estimate the parameters of each LMM, but before examining these we examine the validity of the assumptions behind each LMM. Figure 3.5 displays diagnostic residual plots for the LMM fitted to global data. The plots use raw residuals and indicate that the assumptions behind the LMM are reasonable. The assumption that the residual variation is constant across the GCMs is perhaps questionable (see centre-right panel of Figure 3.5), but overall the LMM appears to be fit for purpose. There is a suggestion that the LMM overestimates the precipitation changes at fitted values under 0 mm day^{-1} (see top-left panel of Figure 3.5), but this is only based on a few data points. The residuals appear to follow a Normal distribution (see bottom panels of Figure 3.5), as assumed by the LMM. Moreover the residuals seem randomly scattered around zero with constant variance when shown against fitted values above 0 mm day^{-1} (top-left panel) and against year (centre-left panel), as assumed. The diagnostic residual plots (not shown) for the LMM fitted to the tropical ocean data show that the validity of the assumptions behind the tropical-ocean LMM is similar to those behind the global LMM.

The fixed parameters for the global data set are shown in the first column of Table 3.2. The estimate of the mean intercept (β_0) highlights that the average (across GCMs) initial decrease (seen in Figure 3.3) in annual-mean precipitation after the forcing is by 0.158 mm day^{-1} in the global data. The estimated average effect (β_1) across GCMs of a 1 K increase in global-mean annual-mean temperature is an increase of 0.080 mm day^{-1} in the global-mean annual-mean precipitation. There is little correlation between the random effects.

	Global	Tropical ocean
β_0	-0.158 mm day^{-1}	-0.255 mm day^{-1}
β_1	0.080 mm day^{-1} K^{-1}	0.132 mm day^{-1} K^{-1}
σ	0.00951 mm day^{-1}	0.0274 mm day^{-1}
sd($\psi^{(0)}$)	0.0261 mm day^{-1}	0.0543 mm day^{-1}
sd($\psi^{(1)}$)	0.00849 mm day^{-1} K^{-1}	0.0148 mm day^{-1} K^{-1}
cor($\psi^{(0)}, \psi^{(1)}$)	0.02	-0.20

Table 3.2: LMM estimates for the global data (first column) and tropical ocean data (second column). These include the standard deviations (sd) of the random effects and the correlation (cor) between them.

The second column of Table 3.2 shows the estimated fixed parameters for the tropical ocean data. There is a greater (0.255 mm day^{-1}) average initial decrease across GCMs in precipitation after the forcing over the tropical ocean than globally. There is also a stronger (0.132 mm day^{-1} per K) relationship between annual-mean temperature change and annual-mean precipitation change over the tropical ocean. This is perhaps because temperature drives convection, and convection is more prominent

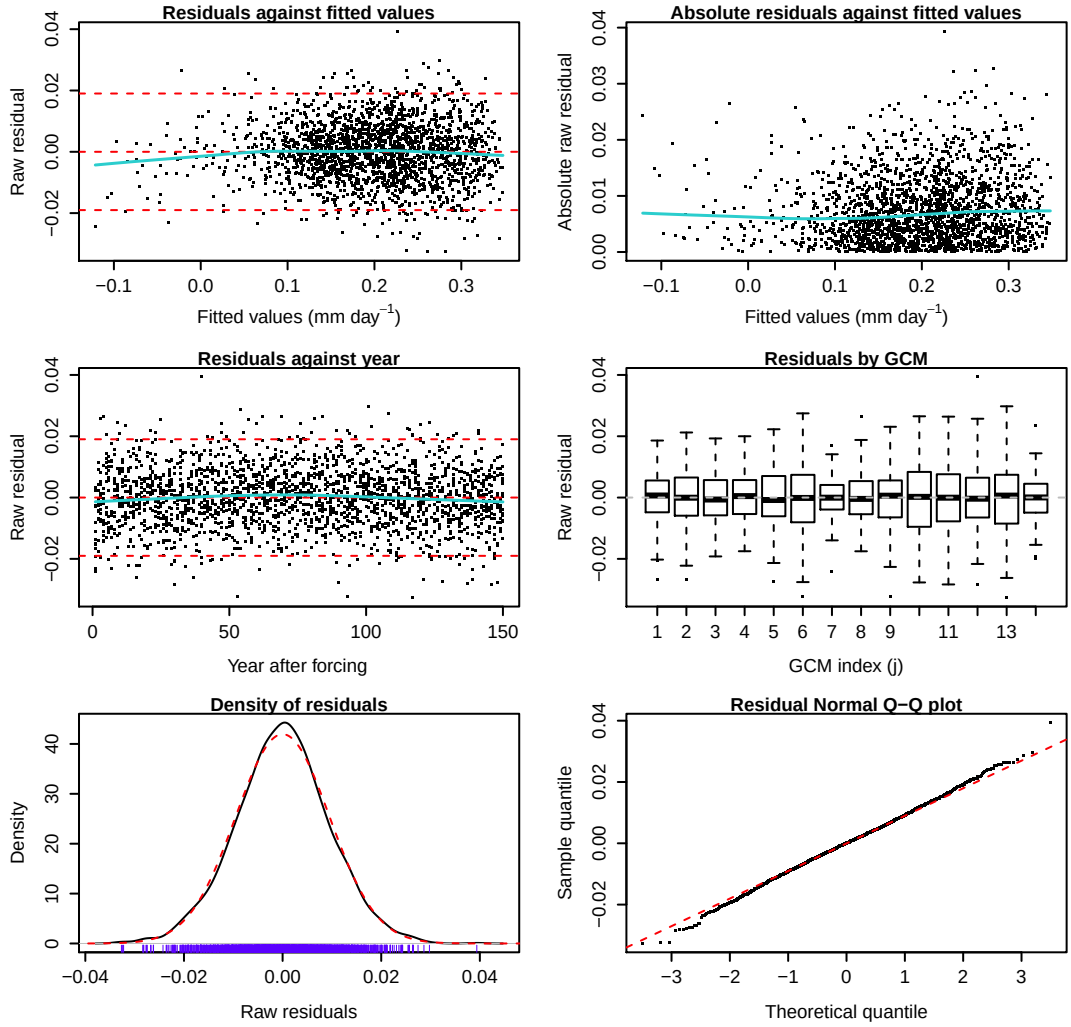

Figure 3.5: Diagnostic residual plots (using raw residuals) for the LMM fitted to global data. The dashed red lines provide an idealized guide, and the solid blue curves are local polynomial smoothers showing any trends where appropriate.

over the tropical ocean (see the introduction to this chapter, Section 3.1). There is more residual variation, more variation across GCMs in the initial decrease in precipitation after the forcing, and more variation across GCMs in the relationship between temperature change and precipitation change over the tropical ocean than globally. However this is perhaps to be expected because the total variation in precipitation change is greater over the tropical ocean (0.0190 mm^2 day^{-2}) than globally (0.00598 mm^2 day^{-2}). Unlike the in global data, there is negative correlation between the random effects over the tropical ocean. This suggests that the variation across GCMs in the effect of temperature change on precipitation compensates for the variation across GCMs in the initial decrease (due to ocean heat uptake) in precipitation after the CO_2 forcing over the tropical ocean, although the correlation is weak.

The fixed parameters show the absolute variation in the residuals and the ran-

dom effects, but to quantify the effect of each source of variation on the spatially-averaged annual-mean precipitation change, the importance measures of Chapter 2 are needed. Before considering these importance measures, we briefly examine the point estimates of the random effects for each GCM (i.e. the intercept and slope estimated for each GCM in the ensemble) and the mean temperature change in response to the CO_2 forcing for each GCM. In particular, the similarity between the different GCMs may be of interest to those who are familiar with the GCMs in the ensemble.

3.5 Similarity between the GCMs

Using the point estimates from the global LMM, the similarity between the GCMs can be examined by comparing

- the change in global-mean annual-mean precipitation (due to ocean heat uptake) immediately after the CO_2 forcing for each GCM (i.e. the variation across the GCMs in the intercept of the LMM, captured by $\psi^{(0)}$),
- the effect of temperature change on the global-mean annual-mean precipitation change in response to the CO_2 forcing for each GCM (i.e. the variation across the GCMs in the slope of the LMM, captured by $\psi^{(1)}$),
- and the distribution of temperature change for each GCM (i.e. the variation across the GCMs in the mean change in temperature, which we denote by γ).

These can also be compared over the tropical ocean.

Figure 3.6 compares $\psi^{(0)}$, $\psi^{(1)}$ and γ for the fourteen GCMs in the ensemble on the global scale. It highlights that bcc-csm1.1, CSIRO-Mk3.6.0, MIROC-ESM and MPI-ESM-LR (coloured red in the figure) have a relatively small initial decrease in global-mean precipitation after the forcing. CCSM4, MIROC5 and NorESM1-M (orange) simulate low temperature change in response to increased CO_2. CanESM2 and HadGEM2-ES (light blue) have a weak temperature-precipitation relationship. GISS-E2-H, IPSL-CM5A-LR and MRI-CGCM3 (dark blue) each seem relatively distinct with respect to their simulation of precipitation and temperature in response to increased CO_2, the latter GCM (MRI-CGCM3) because it is extreme on all three indices: it has a small initial decrease in precipitation after the forcing, a low distribution of temperature change in response to the forcing, and a strong temperature-precipitation relationship.

Figure 3.7 shows the same as Figure 3.6 but for the tropical ocean. This means that not only are $\psi^{(0)}$ and $\psi^{(1)}$ taken from the LMM fitted to the tropical ocean data,

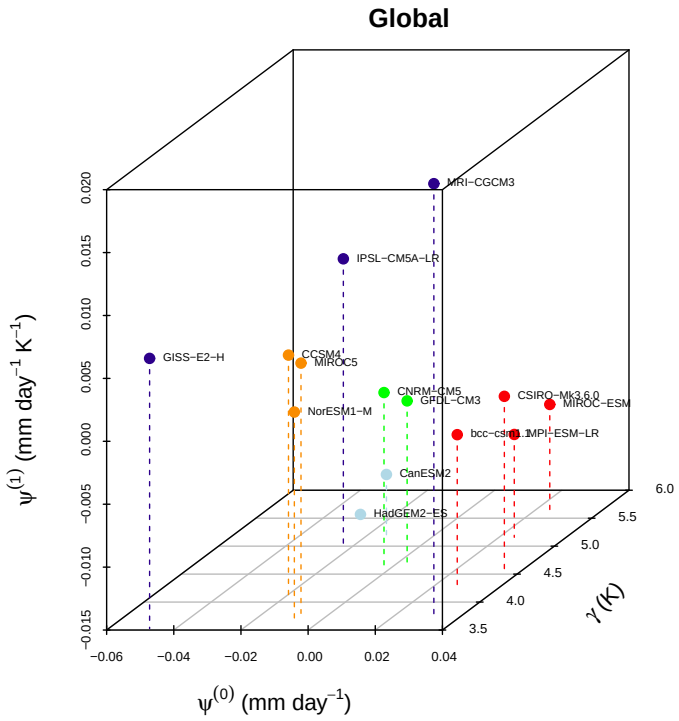

Figure 3.6: GCMs by their change in precipitation immediately after the CO_2 forcing ($\psi^{(0)}$), strength of temperature-precipitation relationship ($\psi^{(1)}$), and mean change in temperature (γ) on the global scale. The former two are expressed as deviations from the ensemble mean. Similar GCMs with respect to $\psi^{(0)}$, $\psi^{(1)}$ and γ are subjectively grouped by colour.

but also γ is the mean temperature change over the tropical ocean in each GCM. The colours of Figure 3.6 are used and suggest that the GCMs have broadly similar characteristics with respect to precipitation and temperature in response to increased CO_2 over the tropical ocean as they do globally. However the GCMs appear more randomly scattered than on the global scale and cluster into less distinct groups.

3.6 Decomposition structure

Following the methodology of Chapter 2, we use our statistical models that relate changes in near-surface air temperature (ΔT) in response to increased CO_2 to changes in spatially-averaged annual-mean precipitation (ΔY) in an ensemble of GCMs and decompose the total variation in the precipitation for each data set. As such, we separate the temperature change into its mean (γ) within each GCM and its temporal deviations ($\Delta T'$) from these means for the global data and for the tropical ocean data, i.e. (including subscripts)

$$\Delta T_{ij} = \gamma_j + \Delta T'_{ij} \, .$$

Therefore the set of inputs to each of our LMMs is

- the variation across GCMs in the intercept, captured by $\psi^{(0)}$,

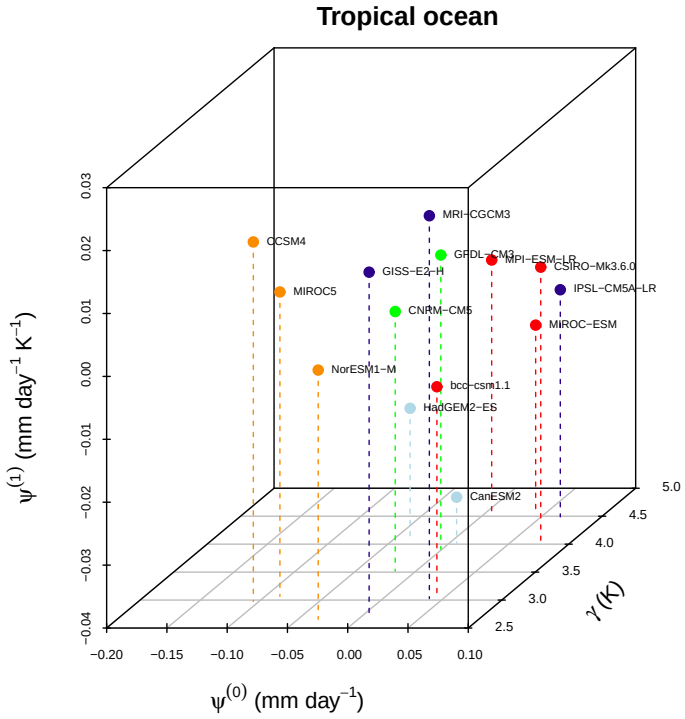

Figure 3.7: As Figure 3.6 but over the tropical ocean. Colours are the same as for Figure 3.6.

- the variation across GCMs in the relationship between temperature change and precipitation change, $\psi^{(1)}$,
- the variation across GCMs in the distribution (mean) of the temperature changes, γ,
- the temporal variation in the temperature changes, $\Delta T'$,

as well as the residual term, ϵ. See Figure 3.8 for a schematic of how these inputs could contribute to the total variation in precipitation change when two GCMs are present. The total variation in precipitation change is represented in the schematic by a sketch of a possible probability density function (PDF) to the right of the figure, $\Delta T'$ is the variation remaining in ΔT after correcting for γ, and ϵ is the variation in ΔY around the curves.

Using the inputs above, we hierarchically decompose the total variation in precipitation change in response to increased CO_2 into the sources of variation shown in Figure 3.9. The predictor is the change in temperature, and so the intercept captures the variation across GCMs in precipitation change with no change in temperature. As such, decomposing the first-order effect of the random effects into their individual first-order effects is physically meaningful and useful in this application. The decomposition shown in Figure 3.9 is a full decomposition of the total variation in the precipitation change because there are only two interactions present between the inputs. The presence of interactions is decided when specifying the statistical

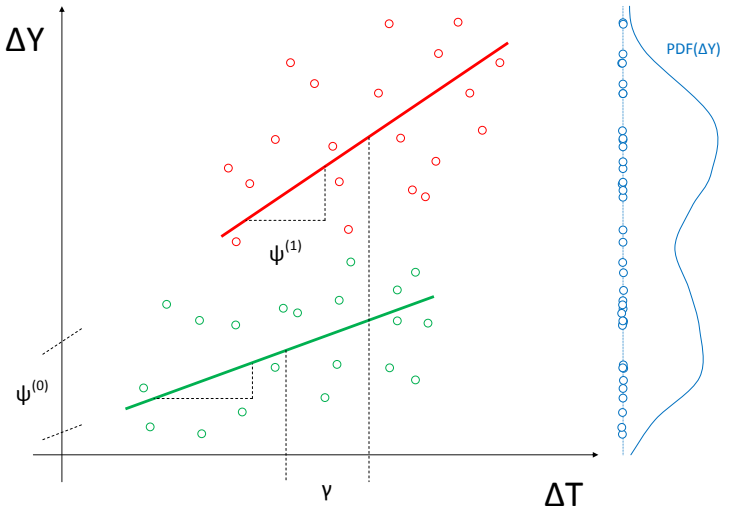

Figure 3.8: Schematic showing the sources of variation in the change in precipitation (ΔY) in response to increased CO_2 captured by a LMM fitted to a two-member (shown in green and red) ensemble of GCMs using the change in temperature (ΔT) in response to the increased CO_2 as a predictor. See text for details.

model, and from the statistical model (recall Section 3.4) it can be seen that only $\Delta T'$ and $\psi^{(1)}$, and γ and $\psi^{(1)}$, combine in multiplicative effects. In the decomposition, the extent to which precipitation change is simulated robustly across GCMs is quantified by $\mathrm{TE}(\gamma, \psi^{(0)}, \psi^{(1)})$, and the dependence of the precipitation change on the temporal variation in temperature change is quantified by $\mathrm{FOE}(\Delta T')$.

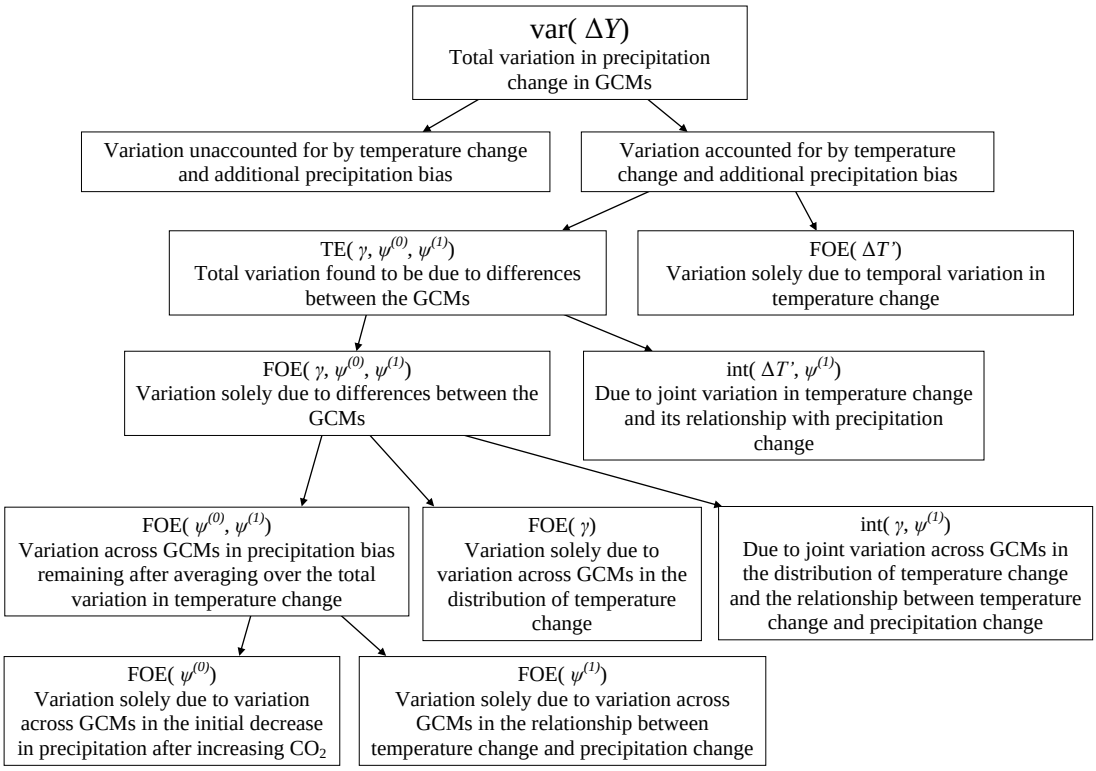

Figure 3.9: Decomposition structure for the change in precipitation (ΔY) in response to increased CO_2 in an ensemble of GCMs. Total effects (TE) and first-order effects (FOE) of inputs (see text) are used, as well as interactions (int) between them.

The variation in precipitation change found to be due to the differences between the GCMs could be because 1) the relationship between temperature change and precipitation change varies across GCMs, 2) the distribution of temperature change varies across GCMs, or 3) the initial decrease in precipitation after increasing CO_2 varies across GCMs. Following the methodology of Chapter 2, we can gain insight into the prominence of these three sources of variation by decomposing the total variation ($\mathrm{TE}(\gamma, \psi^{(0)}, \psi^{(1)})$) in precipitation change found to be due to the differences between the GCMs, as shown in Figure 3.9. Variation across GCMs in the relationship between temperature change and precipitation changes ((1) above) drives the interaction ($\mathrm{int}(\Delta T', \psi^{(1)})$) between the temporal component of the temperature changes and the relationship between the temperature and precipitation changes, and also drives the first-order effect ($\mathrm{FOE}(\psi^{(1)})$) of the relationship. Therefore we can gain insight into the extent to which the relationship between temperature change and precipitation change varies across GCMs by estimating $\mathrm{int}(\Delta T', \psi^{(1)})$ and $\mathrm{FOE}(\psi^{(1)})$. As stated in Section 3.1, Lambert and Webb (2008) found that the relationship between temperature change and global-mean annual-mean precipitation change is dominated by processes that are simulated robustly across GCMs.

Variation across GCMs in the relationship between temperature change and precipitation change ((1) above) and the variation across GCMs in the distribution of temperature change ((2) above) both drive $\mathrm{int}(\gamma, \psi^{(1)})$, whereas variation across GCMs in the distribution of temperature change ((2) above) is the only source driving its first-order effect ($\mathrm{FOE}(\gamma)$). As such, we can gain insight into the extent to which the variation across GCMs in the distribution of temperature change affects the simulation of precipitation by the GCMs by estimating $\mathrm{FOE}(\gamma)$, and to a lesser extent $\mathrm{int}(\gamma, \psi^{(1)})$. Lambert and Webb (2008) found that the distribution of temperature change depends on processes simulated less robustly across GCMs on the global scale than those that drive the relationship between temperature change and precipitation change. The variation across GCMs in the initial decrease in precipitation after increasing CO_2 ((3) above) is isolated by the first-order effect ($\mathrm{FOE}(\psi^{(0)})$) of the intercept.

3.7 Simulation procedure

We use simulation to estimate the terms in Figure 3.9, and follow the simulation procedure recommended in Subsection 2.2.2 and Section 2.4. Following this simulation procedure, independence is assumed between the input ($\Delta T'$) that captures the temporal variation in the predictor and the inputs ($\psi^{(0)}$, $\psi^{(1)}$ and γ) that capture the variation across GCMs. The box plots of the top panels of Figure 3.4 (page 97)

demonstrate that there is some variation across the GCMs in the shape of the distribution of the temperature changes (not just their means) in both data sets, with relatively small variance of ΔT within a GCM if the mean of ΔT is low in that GCM. This indicates that assuming independence between $\Delta T'$ and the other inputs is a simplification in this application. However, as discussed in Subsection 2.2.2, $\Delta T'$ is centred on zero within each GCM (by definition) and so there is no first-order dependence between $\Delta T'$ and any other input. Therefore we make the simplification of assuming independence between $\Delta T'$ and the other inputs because accounting for variation across the GCMs in the shape of the distribution of the temperature changes using the methods discussed in Section 2.4 would substantially complicate the decomposition.

Following the methods suggested in Section 2.4 for each data set, samples for $\Delta T'$ are taken from its empirical distribution, and $\psi^{(0)}$, $\psi^{(1)}$ and γ are sampled using a multivariate Gaussian copula with the empirical marginal distribution of γ. The marginal distributions of the random effects are the Gaussian distributions taken from the fitted LMM for each data set. The dependence structure for the multivariate Gaussian copula is determined by the estimates from the LMM of the variance components of the random effects, and from the empirical data for the variation in γ. The correlations between the random effects and γ are estimated by using the point estimates of the random effects for the fourteen GCMs in the ensemble, along with the empirical values of γ. These are shown in Table 3.3.

	Global	Tropical ocean
$\text{cor}(\gamma, \psi^{(0)})$	0.15	0.46
$\text{cor}(\gamma, \psi^{(1)})$	-0.71	-0.52

Table 3.3: Correlation between the random effects ($\psi^{(0)}$ and $\psi^{(1)}$) and the mean change in temperature (γ) in response to increased CO_2 for the global data (first column) and tropical ocean data (second column).

Table 3.3 shows that there is strong correlation between the random effects and γ for both analyses. The strong negative correlation (-0.71 globally and -0.52 over the tropical ocean) between $\psi^{(1)}$ and γ shows that GCMs that simulate high temperature changes tend to simulate a weak (positive) relationship between temperature change and precipitation change (and vice versa). This suggests that the physical constraints on temperature and its effect on precipitation prevent both from being high in a GCM. As outlined in the introduction to this chapter (Section 3.1), the physical constraints on temperature and precipitation are stronger globally, and the negative correlation between $\psi^{(1)}$ and γ is indeed stronger in the global data than

over the tropical ocean. Furthermore, the positive correlation (0.15 globally and 0.46 over the tropical ocean) between $\psi^{(0)}$ and γ highlights that GCMs that simulate high temperature changes tend to also simulate higher than average initial precipitation change after the forcing (and vice versa). This correlation is weak globally but stronger over the tropical ocean.

3.8 Results

The black values of Figure 3.10 show the decomposition of the total variation in global-mean annual-mean precipitation change in response to increased CO_2. The results are based on 16 000 samples and size 2000 ($R = 16\,000$ and $r = 2000$ using the notation of Section 2.3). The total variation in the global-mean precipitation change is estimated (0.00681 mm^2 day^{-2}) to be close enough to the empirical (0.00598 mm^2 day^{-2}) to indicate that the statistical modelling captures the data well. Of the total variation in the global-mean precipitation, 99% is accounted for by the temperature change and additional precipitation bias. Much of this is solely due to the temporal variation in temperature change in response to the forcing. Through the Clausius-Clapeyron relation, temperature is known to be a driver of precipitation on the global scale, and FOE($\Delta T'$) shows that 53% of the variation in the precipitation change is solely due to the temporal variation in temperature change in the ensemble.

From TE($\gamma, \psi^{(0)}, \psi^{(1)}$) in Figure 3.10, we find the simulation of the change in global-mean precipitation in response to increased CO_2 to be highly sensitivity to the choice of GCM. Despite global-mean precipitation being constrained by the tropospheric energy budget and the Clausius-Clapeyron relation (recall from Section 3.1), we find that 46% of the total variation in global-mean annual-mean precipitation change is due to the differences between the GCMs.

As a result of the strong dependence between inputs, the total variation in precipitation change that is found to be due to the differences between the GCMs cannot be decomposed orthogonally. Despite this, it is clear that there is little variation across GCMs in the relationship between temperature change and precipitation change. This can be seen because the interaction (int($\Delta T', \psi^{(1)}$)) between this relationship and the temporal component of temperature change (which is unaffected by the dependence between inputs) only accounts for 1% of the total variation in precipitation change. This is despite the importance of the temporal variation in the temperature seen above by FOE($\Delta T'$). Moreover, the first-order effect (FOE($\psi^{(1)}$)) of the relationship between temperature change and precipitation change is estimated to be 1% (albeit affected by the dependence between inputs). This implies that the

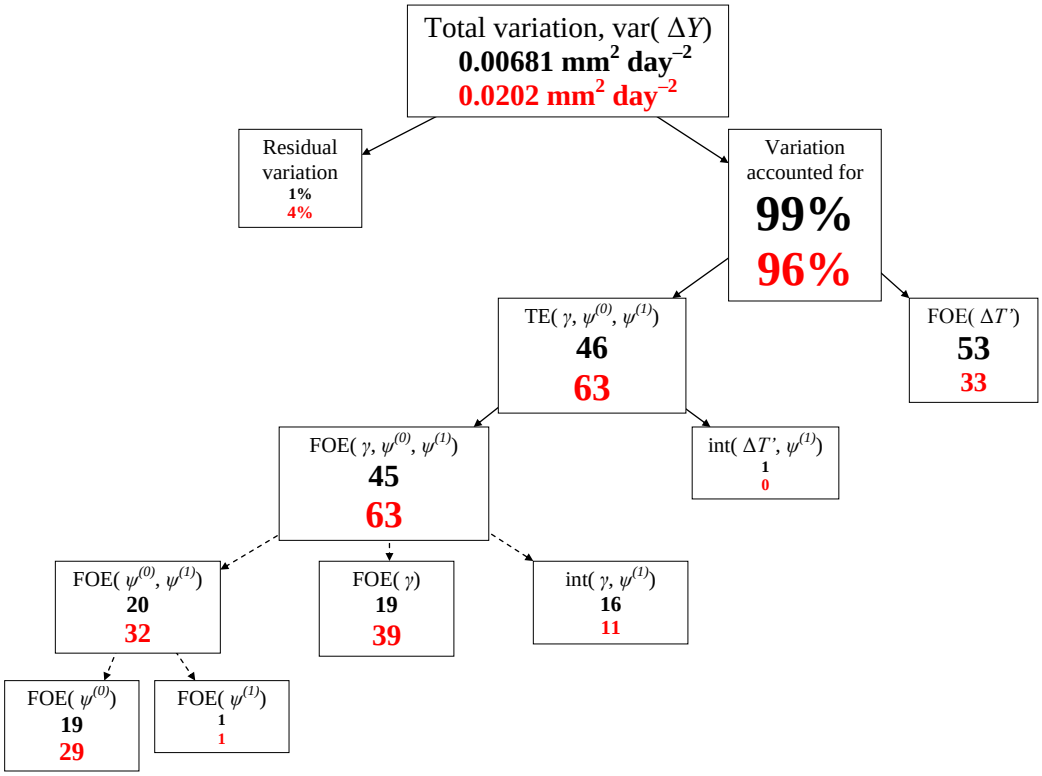

Figure 3.10: Decomposition of the total variation in the change in annual-mean global-mean (black) and tropical ocean-mean (red) precipitation (ΔY) in response to increased CO_2. Each total effect (TE), first-order effect (FOE) and interaction (int) is expressed as a percentage of var(ΔY) (rounded to integers). Dashed lines indicate non-orthogonality due to dependence between inputs. The empirical variation in ΔY is 0.00598 mm^2 day^{-2} globally and 0.0190 mm^2 day^{-2} over the tropical ocean.

relationship between temperature change and precipitation change is represented robustly across GCMs. The estimate (19%) of FOE($\psi^{(0)}$) shown in Figure 3.10 suggests that there is important variation across GCMs in the decrease (due to ocean heat uptake) in precipitation immediately after a CO_2 forcing.

In addition to FOE($\psi^{(0)}$) and FOE($\psi^{(1)}$), the estimated variation solely due to the variation (FOE(γ)) across GCMs in the distribution of temperature change, and the interaction (int($\gamma, \psi^{(1)}$)) between the distribution of temperature change and the relationship between the temperature and precipitation changes, are also affected by the dependence between the inputs. Despite this, there seems substantially more variation due to variation across GCMs in the distribution of temperature change than there is due to variation across GCMs in the relationship between temperature change and precipitation change. This can be seen because FOE(γ) is estimated to be 19% and int($\gamma, \psi^{(1)}$) as 16%, whereas (as seen above) FOE($\psi^{(1)}$) and int($\Delta T', \psi^{(1)}$) are both estimated to be only 1%. Therefore it is clear that there is substantial variation across GCMs in the distribution of temperature change, and that this substantially impacts precipitation. This variation across GCMs in the distribution of global temperature change is despite the physical constraint of the

global energy conservation.

The red values of Figure 3.10 show the decomposition of the total variation in tropical ocean-mean annual-mean precipitation change in response to increased CO_2. There is more variation in the precipitation averaged over the tropical ocean than averaged globally, and as with the decomposition of the global-mean precipitation, the total variation in the precipitation averaged over the tropical ocean is estimated (0.0202 mm^2 day^{-2}) to be relatively close to the empirical (0.0190 mm^2 day^{-2}). Proportionately less of this variation is accounted for over the tropical ocean (96%) than globally (99%). This is despite the greater prominence of convection over the tropics, which is driven by temperature, and so is likely to be as a result of the weaker physical constraints over the tropics leading to greater changes in precipitation independent of temperature in GCMs.

The weaker physical constraints and greater prominence of convection over the tropical ocean than globally result in greater disagreement between GCMs in their simulation of tropical ocean-mean precipitation change in response to increased CO_2. Of the variation in annual-mean precipitation change over the tropical ocean, in total 63% ($\text{TE}(\gamma, \psi^{(0)}, \psi^{(1)})$) is due to the differences between GCMs. Given that 4% is residual variation, this leaves only 33% of the variation in the precipitation change accounted for by the temporal variation ($\text{FOE}(\Delta T')$) in temperature change over the tropical ocean. Therefore we find that not only is the global-mean precipitation change in response to increased CO_2 highly sensitive to the choice of GCM, but that the choice of GCM can be substantially more important when analysing precipitation on smaller spatial scales.

As with the global data, the relationship between temperature change and precipitation change appears to be robust ($\text{int}(\Delta T', \psi^{(1)}) = 0\%$ and $\text{FOE}(\psi^{(1)}) = 1\%$) across the GCMs over the tropical ocean. This is in contrast to the effect on the precipitation of the variation across GCMs in the distribution of the temperature changes ($\text{FOE}(\gamma) = 39\%$). It is difficult to compare the rest of the decomposition over the tropical ocean to that globally because of the different dependence structure between the inputs, but there appears to be greater variation ($\text{FOE}(\psi^{(0)}) = 29\%$) across GCMs in the extent to which ocean heat uptake reduces precipitation immediately after the increased CO_2 over the tropical ocean.

3.9 Summary

We find large disagreement between GCMs in their simulation of the change in global-mean precipitation in response to increased CO_2. We find that 46% of the

total variation in global-mean annual-mean precipitation change in an ensemble of GCMs is due to the differences between the GCMs. This is found to be largely because the distribution of temperature change is not simulated robustly across the GCMs, and the precipitation depends strongly on temperature. We also found variation across GCMs in the initial decrease in precipitation (due to ocean heat uptake) immediately after the CO_2 forcing. The relationship between temperature change and precipitation change is represented relatively robustly across GCMs.

The sensitivity of the global-mean precipitation change to choice of GCM is despite the tropospheric energy budget constraining global-mean precipitation. Averaging the changes in precipitation over smaller regions weakens these physical constraints, and so the precipitation can be simulated even less robustly across GCMs than is the case globally. In addition to physical constraints, the lack of robustness across GCMs may also depend on the prominence of convective precipitation, which is difficult to represent in GCMs, and on the extent to which the surface of the region is ocean. As a result, we find that 63% of the total variation in tropical ocean-mean annual-mean precipitation change in our ensemble of GCMs is due to the differences between the GCMs. Therefore on this spatial and temporal scale, the precipitation depends on the choice of GCM substantially more than on the temporal variation within each GCM.

3.10 Discussion

Temperature dictates the capacity of air to hold moisture and drives convection, and so the dependence of precipitation on temperature within GCMs that we have found is not surprising. The importance of the temporal variation in temperature change can, however, depend on the strength of the time trend exhibited by the changes in temperature. If the time trend in the temperature changes was weaker, then the temporal variation would reduce in both temperature and precipitation, and so the importance of FOE($\Delta T'$) would reduce. To see this, recall the time series of Figure 3.2 on page 93: the time trends are greatest soon after the forcing, resulting in the temporal variation in the left half of the time series being more important than those in the right halves. Reducing FOE($\Delta T'$) would increase the prominence of the differences between GCMs, i.e. it would increase TE($\gamma, \psi^{(0)}, \psi^{(1)}$). The strength of the time trends can be made weaker by either considering years that are further in time from the forcing, or by weakening the strength of the forcing (e.g. by doubling CO_2 instead of quadrupling it). Therefore the relative sizes of FOE($\Delta T'$) and TE($\gamma, \psi^{(0)}, \psi^{(1)}$) can depend on the temporal period after the forcing, and on the size of the forcing. Note, however, that these are only issues in this chapter because of the type of data used, i.e. because the data are changes in response to a

forcing, whereas in the other application chapters we use data from simulated real world years (i.e. with dates, e.g. 1990), and so we simply use the years in which we are interested. Such simulated years are forced by observed levels of greenhouse gases, and so the change in the forcing is more gradual and resulting trends are generally less pronounced.

Despite the dependence of FOE($\Delta T'$) and TE($\gamma, \psi^{(0)}, \psi^{(1)}$) on the temporal period after the forcing and the size of the forcing, the lack of robustness across GCMs that we found in the change in precipitation in the 150 years after quadrupling CO_2 is perhaps concerning. We agree with Lambert and Webb (2008) that the main source of this uncertainty is the disagreement between GCMs in their simulated distribution of temperature change, and not the effect of the changes in temperature on precipitation. Therefore if the goal is to develop GCMs such that they simulate precipitation more robustly and provide more consistent projections, then the focus should be on reducing the uncertainty in the distribution of temperature change.

Given the importance of the distribution of temperature change to the precipitation change, it is worth focusing on which GCMs simulate high temperature changes in response to the increased CO_2, which simulate low temperature changes, and which simulate average temperature changes. Re-visiting the top-left panel of Figure 3.2 (page 93), the top-left panel of Figure 3.4 (page 97) and Figure 3.6 (page 101), MIROC-ESM and HadGEM2-ES simulate high temperature changes, GISS-E2-H, NorESM1-M, MIROC5 and MRI-CGCM3 simulate low temperature changes, and CSIRO-Mk3.6.0, CNRM-CM5 and GFDL-CM3 simulate average temperature changes. Lambert and Webb (2008) cite the uncertainty in how to represent shortwave cloud feedbacks in GCMs as a driver of the variation across the GCMs in the distribution of temperature change. As seen in Subsection 1.4.2, studies have found that the entrainment coefficient and ice-fall speed parameter influence the distribution of temperature change in climate models (Stainford et al., 2005; Sanderson et al., 2008b; Fowler et al., 2010). Future work, therefore, could be to quantify the extent to which these and other differences between the GCMs contribute to the variation in how they represent shortwave cloud feedbacks.

With respect to developing the statistical modelling used in this analysis, one possibility would be to more formally model the time dependence in the data. For example this could be done by including a time predictor. Given that the changes in temperature are strongly dependent on time, however, any time predictor would be so strongly correlated with the change in temperature that disentangling their individual effects could be difficult. Other options for modelling time dependence include using an autoregressive term or perhaps lagged values of the change in tem-

perature. Although these could improve the modelling, we do not feel that they are necessary to obtain useful results.

4 Summer wet-day convective precipitation

Aim: To quantify the extent to which, and understand why, convective precipitation varies across and within perturbed physics ensembles of climate models.

4.1 Introduction

As discussed in Section 1.6, quantifying the extent to which climate models agree in their simulation of precipitation is important when judging the clarity of projections, and can affect our confidence in the projections. Moreover, examining why the simulation of precipitation varies across climate models is important when developing the models, and examining why the precipitation varies within climate models can aid understanding of the driving physical processes of precipitation in the real world. Precipitation can be classified as either large scale (dynamical) or convective (see Section 1.3), and this results in climate models often including large-scale and convective precipitation as separate variables in their output (see Subsection 1.4.1). As outlined in the introduction to Chapter 3 (Section 3.1), convective precipitation generally occurs on finer spatial and temporal scales than large-scale precipitation, and the fine spatial scale leads to the parameterization of convective precipitation in climate models (also see Subsection 1.4.1). Therefore we examine the extent to which perturbing the parameters of a regional climate model (RCM) affects its simulation of convective precipitation. We attempt to do this in a more general sense than restricting our analysis to the mean rate of precipitation within each ensemble member, and so we examine the representation by the climate model of the driving physical process of convective precipitation on a local spatial (the grid-point level of an RCM) and temporal (daily totals) scale.

As seen in Section 1.3, in the real world convective precipitation is driven by convective buoyancy of moist air in the atmosphere. Convective buoyancy occurs when relatively warm air is below cooler air such that the lapse rate (or vertical temperature gradient through the atmosphere) is great enough to trigger convection. This convection causes the air to rise, and so cools adiabatically and its capacity to hold moisture is reduced. If the capacity of the air to hold moisture (i.e. its saturation specific humidity) is reduced to lower than the amount of moisture that it is holding, such that its relative humidity reaches 100%, then the moisture condenses and convective precipitation occurs. Therefore we capture the representation of this driving physical process of convective precipitation in a perturbed physics ensemble using a generalized linear mixed model (GLMM) and apply the methodology of Chapter 2. In doing so we decompose the total variation in convective precipitation in the en-

semble, and attempt to quantify the robustness of the simulation of the convective precipitation events.

Variation across the members of an ensemble of climate models in the simulation of convective precipitation may or may not result in variation across ensemble members in their mean rate of convective precipitation (i.e. precipitation bias) (see Section 1.8). This is because given atmospheric conditions could lead to different amounts of convective precipitation within each ensemble member, and given atmospheric conditions could occur at different rates. This could result in variation across ensemble members in the simulation of convective precipitation that is not detectable from comparing the mean convective precipitation rate within each ensemble member. Therefore we use our GLMM to relate lapse rate, saturation specific humidity and relative humidity to the convective precipitation in a perturbed physics ensemble to capture the driving physical process of convective precipitation described above.

Different sets of predictors than lapse rate, saturation specific humidity and relative humidity could be used, for example the inputs to the convection scheme of the RCM (i.e. the variables that dictate the amount of convective precipitation outputted by the convection scheme, recall background on convection schemes in Subsection 1.4.1). Using such predictors would perhaps allow more of the variation in the convective precipitation to be accounted for by the GLMM, but some of the variation would remain unaccounted for because of the temporal resolution of the data, and using all the variables at all relevant atmospheric levels could lead to a complicated set of predictors. In contrast, using lapse rate, saturation specific humidity and relative humidity as predictors is simple enough to relate to the physics of the real world. It should also be noted that saturation specific humidity (i.e. the amount of moisture that the air is able to hold) is dictated by the temperature of the air (recall from Section 1.3) and so temperature could be used instead to capture the effect of saturation specific humidity. However given that lapse rate is included as a predictor, the dominant remaining effect on convective precipitation of temperature is likely to be its effect on the capacity of air on hold moisture (i.e. its effect on saturation specific humidity). As such, we choose to use saturation specific humidity instead of temperature as a predictor because saturation specific humidity directly measures the represented physical phenomenon that we are attempting to capture with the predictor, and so its effect on the convective precipitation can be interpreted more directly.

Using lapse rate, saturation specific humidity and relative humidity as predictors allows their relative contributions to the variation in convective precipitation events

to be assessed. Therefore in this chapter we examine whether or not there is a dominant driver of convective precipitation events, i.e. lapse rate or the humidities. In addition to this, if the variation in lapse rate, saturation specific humidity and relative humidity are able to account for enough of the variation in the convective precipitation, then using these predictors will allow investigation into why the simulation of convective precipitation varies across ensemble members. More specifically, we assess the extent to which the uncertainty in RCM parameters affects the relationship between the represented atmospheric conditions captured by the predictors and the precipitation, as well as extent to which the uncertainty affects the frequency distribution of the driving atmospheric conditions of convective precipitation. Furthermore, we aim to quantify the overall extent to which the simulation of convective precipitation events is robust to the uncertainty in the RCM parameters by aiming to use the predictors to capture a large proportion of the variation in the precipitation and quantifying the total contribution to this of the differences between the ensemble members.

Importance measures not only allow the investigations described above, but do so by providing quantitative measures. As such, in this chapter we estimate the importance measures at each grid point over a region of interest (outlined in the next section), produce maps of the results, and examine the spatial patterns exhibited. Determining the causes of such spatial patterns can further reveal why convective precipitation varies across and within the ensemble members. This amount of detail, clarity and quantification would be difficult to achieve using simpler methods, such as by examining exploratory plots of relationships between predictors and the precipitation at different grid points.

4.2 Data

To retain the direct physical link between the predictor variables and convective precipitation, we use data at the grid-point level (treating each grid point separately) and on the daily scale, such that lapse rate (L), saturation specific humidity (q_s) and relative humidity (H, to shorten RH that was used to denote relative humidity in Section 1.3) are taken as daily means, and are related to daily-total convective precipitation. To summarize the lapse rate we take the difference between the atmospheric temperature at 850 hPa and that at 500 hPa, such that the higher the values of our lapse rate, the greater the convective instability of the atmosphere. Saturation specific humidity and relative humidity are taken at 850 hPa. This atmospheric pressure level can summarize the atmospheric conditions because it is usually approximately at 1.5 km altitude, and so close to, but generally above, the boundary layer. As a result the 850 hPa atmospheric pressure level can summa-

rize the free atmosphere without being overly sensitive to surface conditions and parameterizations.

We choose to analyse the convective precipitation over Britain. Over Britain, convective precipitation is most prominent in summer and across its southern half, and even then there are many days in which little (or no) convective precipitation falls. We therefore focus on summer (June, July and August) days of convective precipitation greater than 0.1 mm at the grid-point level across southern Britain. We define a wet day to have greater than 0.1 mm of convective precipitation because this threshold has often been used by previous studies (see Bärring et al., 2006, who state that 0.1 mm is the standard observational threshold). As a result, our response variable (Y) is the convective precipitation excess over this low threshold of 0.1 mm.

The RCM perturbed physics ensemble that we use is of HadRM3 (Met Office Hadley Centre, 2008). HadRM3 provides gridded data with a spatial resolution of 25 km, and so convective precipitation is parameterized. The ensemble is from the HadRM3-PPE-UK experiment (again, see Met Office Hadley Centre, 2008) at the UK Met Office Hadley Centre, which covered the wider European area after dynamically downscaling a general circulation model (GCM), and spanned the period of 1950-2099 (using historical greenhouse gas forcings followed by the medium emission scenario SRESA1B; Nakicenovic and Swart, 2000). The British Isles region is extracted and used for the UK-Climate Projections project (UKCP) (Murphy et al., 2009), from which we extract the region of southern Britain. We also reduce the time period by extracting the 30-year period 1960-1989. This is because this period is short enough not to be overly affected by long-term time trends in the data, and is before the unobserved future emissions scenario begins. For this analysis we are interested in quantifying and understanding the simulation of convective precipitation in the recent climate with observed forcings, rather than also including uncertainty in the future climate.

The ensemble consists of eleven members, one control run and ten runs with the values for the parameters of the subgrid-scale parameterizations of HadRM3 perturbed to capture some of the uncertainty in their optimal values. Our ensemble therefore consists of eleven time series, one per member, spanning the same time period. The parameters that are perturbed in the ensemble are shown in Table 4.1. Note that although the ensemble samples the uncertainty in the physical parameters of the RCM (HadRM3), since the boundary conditions of the area covered by the RCM are driven by a GCM, any error relative to the real world (e.g. an unrealistic storm track across the Atlantic Ocean) in this GCM would also impact on the RCM

output. Therefore given that the RCM is not driven by reanalysis data (i.e. observations dynamically interpolated by a GCM), it is possible that the relationships between the predictors and convective precipitation could depend on the GCM that drives the boundary conditions. We do not investigate this, however, because doing so would require data from HadRM3 with boundary conditions driven by different GCMs, and we do not have access to such data.

4.3 Exploratory data analysis

We now plot the data to identify the sources of variation that may contribute to the variation in convective precipitation in the ensemble. As an illustration, we do this at a single grid point over Birmingham in the English Midlands. Figure 4.1 shows the relationship between lapse rate and convective precipitation on wet days in each of the eleven ensemble members. There is a positive relationship, as expected, in all ensemble members at lapse rates less than 28 K. At lapse rates greater than 28 K, however, the relationship is negative in most ensemble members. This is because these days of high lapse rates are often associated (not shown) with high (surface) temperatures and low relative humidity, and emphasizes that although lapse rate is a main driver of convective precipitation, humidity is also important in determining the daily totals of the convective precipitation. There seems some variation across ensemble members in the relationship between lapse rate and convective precipitation, but not a great amount compared to the total daily variability in the precipitation.

Figure 4.2 suggests that there is little relationship between saturation specific humidity and convective precipitation, although days of heavy convective precipitation do not tend to occur at low saturation specific humidity. The weak relationship between saturation specific humidity and convective precipitation is because saturation specific humidity can have conflicting effects on precipitation. High saturation specific humidity allows the air to hold a high amount of moisture, which can lead to a high rate of precipitation. However the high capacity of the air to hold moisture also means that a greater amount of moisture is required for it to condense and fall as precipitation. Despite its weak univariate effect on convective precipitation, saturation specific humidity is worth including as a predictor of the convective precipitation because we use it along with relative humidity to capture more detail in the atmospheric conditions than if we used only one of these humidities. When saturation specific humidity and relative humidity are simultaneously high, then one would expect convective instability to lead to a high rate of precipitation, but this is not necessarily the case when only one of the humidity variables is high. There seems little variation across ensemble members in the relationship between

	Parameter	Description
Large-scale cloud		
1	RH_{crit}	Threshold of relative humidity for cloud formation
2	Ice-fall speed, V_{f1}	Ice particle fall speed
3	C_w	Cloud droplet to rain conversion threshold over land
4	C_w	Cloud droplet to rain conversion threshold over sea
5	C_t	Cloud droplet to rain conversion rate
6	Cloud fraction at grid-box saturation boundary layer	Cloud cover calculation
7	Cloud fraction at grid-box saturation free troposphere value	
8	Flow-dependent RH_{crit} scheme on/off	Parameterization of RH_{crit} in terms of local variance of cloud water
Convection		
9	Entrainment coefficient	Rate of mixing between environmental air and convective plume
10	Timescale for destruction of CAPE	Intensity of convective mass flux
11	Convective updraught factor	Fraction of convective cloud in which updraught occurs
12	Convective anvil shape factor	Shape of convective cloud
Radiation		
13	Ice particle size (μm)	Effective radius of cloud ice spheres
14	Non-spherical ice particles stratospheric shortwave	Accounts for effect of non-spherical ice particles
15	Non-spherical ice particles stratospheric longwave	
16	Non-spherical ice particles convective shortwave	
17	Non-spherical ice particles convective longwave	
18	Shortwave water vapour continuum absorption on/off	Accounts for shortwave absorption due to the self-broadened continuum of water vapour
Sea ice		
19	Sea ice albedo	Dependence of sea ice albedo on temperature
20	Ocean-ice diffusion coefficient	Ocean to heat transfer
21	Sea ice minimum temperature	
Boundary layer and surface processes		
22	Boundary layer flux profile parameter, G_0	Functions used to determine stability dependence of turbulent mixing coefficients
23	Asymptotic neutral mixing length parameter, λ	Calculation of turbulent mixing coefficients
24	Charnock constant	Roughness lengths and surface fluxes over sea
25	Free convective roughness length over sea	Surface fluxes over tropical oceans
26	Roughness length, z_0	Surfaces fluxes over areas containing forest
27	Number of soil levels accessible for transpiration	Root depths
28	Surface-canopy decoupling scheme on/off	Accounts for effect of vegetation canopy on surface energy budget
29	Stomatal conductance response to ΔCO_2 on/off	Variance of stomatal conductance with carbon dioxide concentration
Dynamics		
30	Dynamical diffusion e-folding time	Diffusion coefficients for heat, momentum and moisture
31	Surface gravity wave constant	Magnitude of hydrostatic gravity wave stress
32	Trapped lee wave constant	Magnitude of non-hydrostatic gravity wave stress
33	Gravity wave drag start level	Lowest model level at which drag is applied

Table 4.1: List of perturbed parameters of HadRM3 and their descriptions. Descriptions are taken from Barnett et al. (2006).

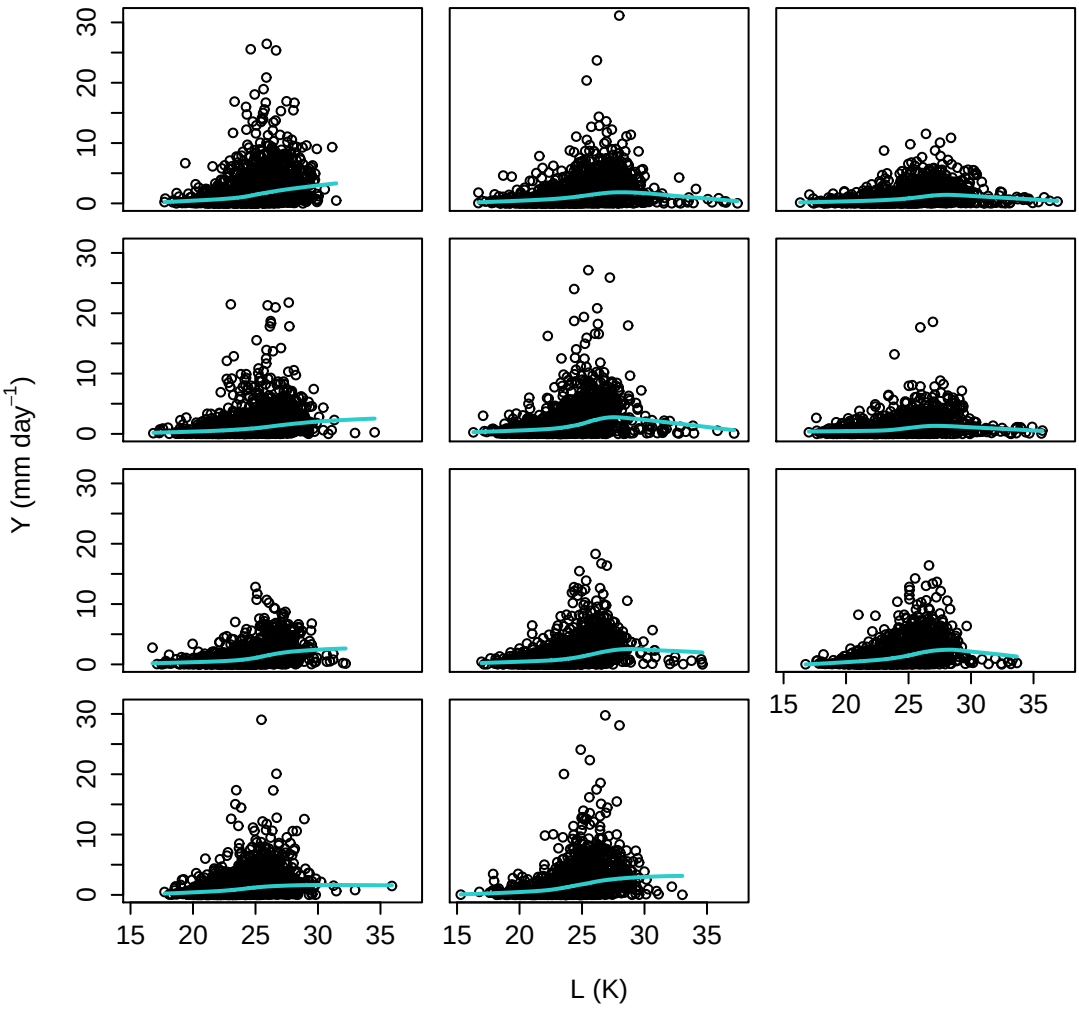

Figure 4.1: Daily-total convective precipitation (excess over 0.1 mm) (Y) against daily-mean lapse rate (L) on wet days in each ensemble member at a single grid point over Birmingham. Curves are local polynomial smoothers. The control run is top left.

saturation specific humidity and convective precipitation.

As with saturation specific humidity, there appears to be little variation across ensemble members in the relationship between relative humidity and convective precipitation compared to the total variation in the precipitation (see Figure 4.3). Relative humidity appears to have a positive effect on convective precipitation at relative humidity greater than 70% in some ensemble members but little effect in others. There tends to be little convective precipitation on days of low relative humidity in all ensemble members. The weak univariate relationship between relative humidity and convective precipitation at relative humidity greater than 70% is due to a negative association between relative humidity and the other predictors. This negative association is because warm conditions are associated with high lapse rate and saturation specific humidity, but low relative humidity, and vice versa for cold conditions. See Figure 4.4 for a three-dimensional plot showing convective precip-

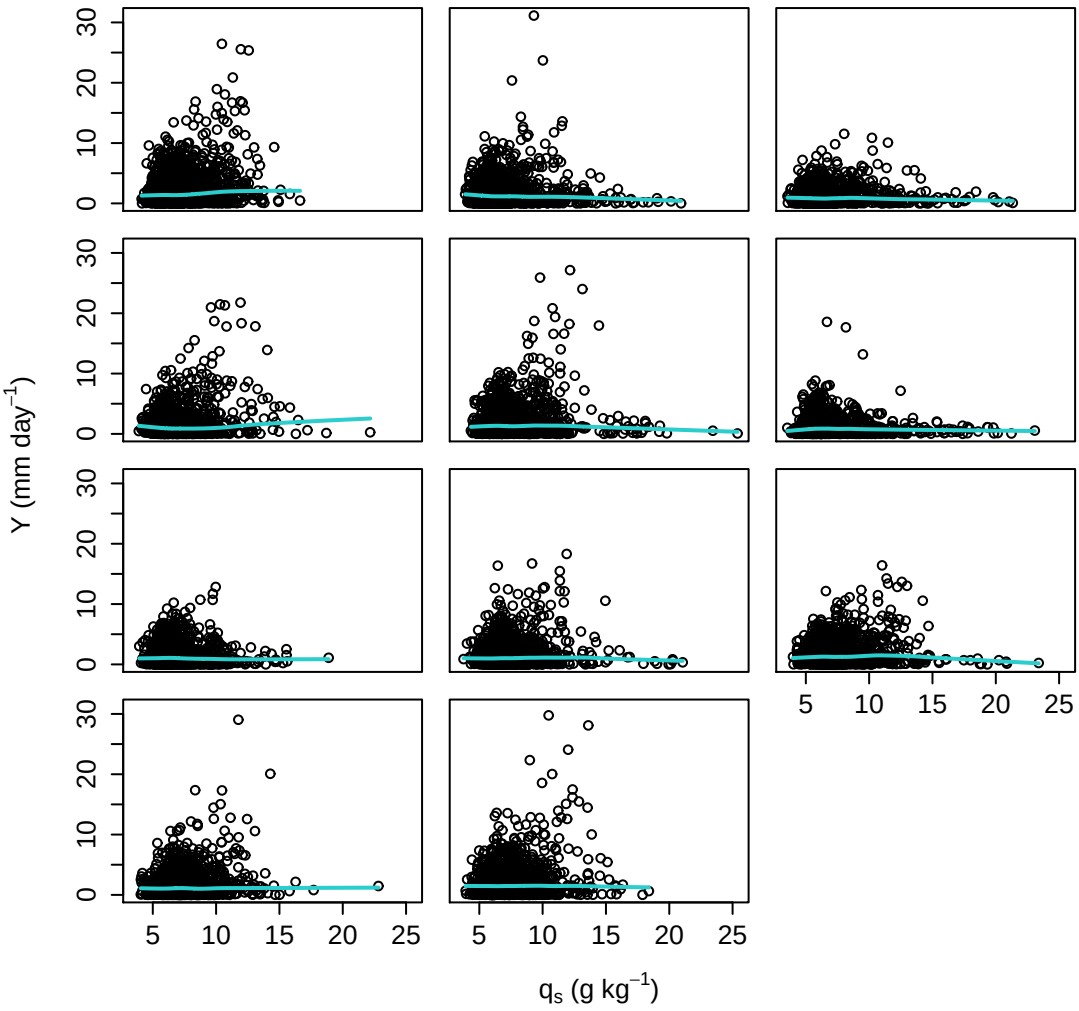

Figure 4.2: Daily-total convective precipitation (excess over 0.1 mm) (Y) against daily-mean saturation specific humidity (q_s) on wet days in each ensemble member at a single grid point over Birmingham. Curves are local polynomial smoothers. The control run is top left.

itation against both saturation specific humdiity and relative humidity. From this figure the negative relationship can be seen between the two humidities, but it also highlights that days of heavy convective precipiation tend to occur when relative humidity is less than 100%, which is partly a result of greater saturation specific humidity.

The top-left panel of Figure 4.5 indicates that there is little variation across ensemble members in the distribution of convective precipitation compared to its temporal variation within each ensemble member. The extent to which the distribution of convective precipitation is positively skewed appears to vary across ensemble members, which is likely to be partly due to the differences between the members and partly due to the temporal variation, but the average convective precipitation rate appears to be similar within each ensemble member. There also appears to be little variation across ensemble members in the distributions of lapse rate (top-right panel), sat-

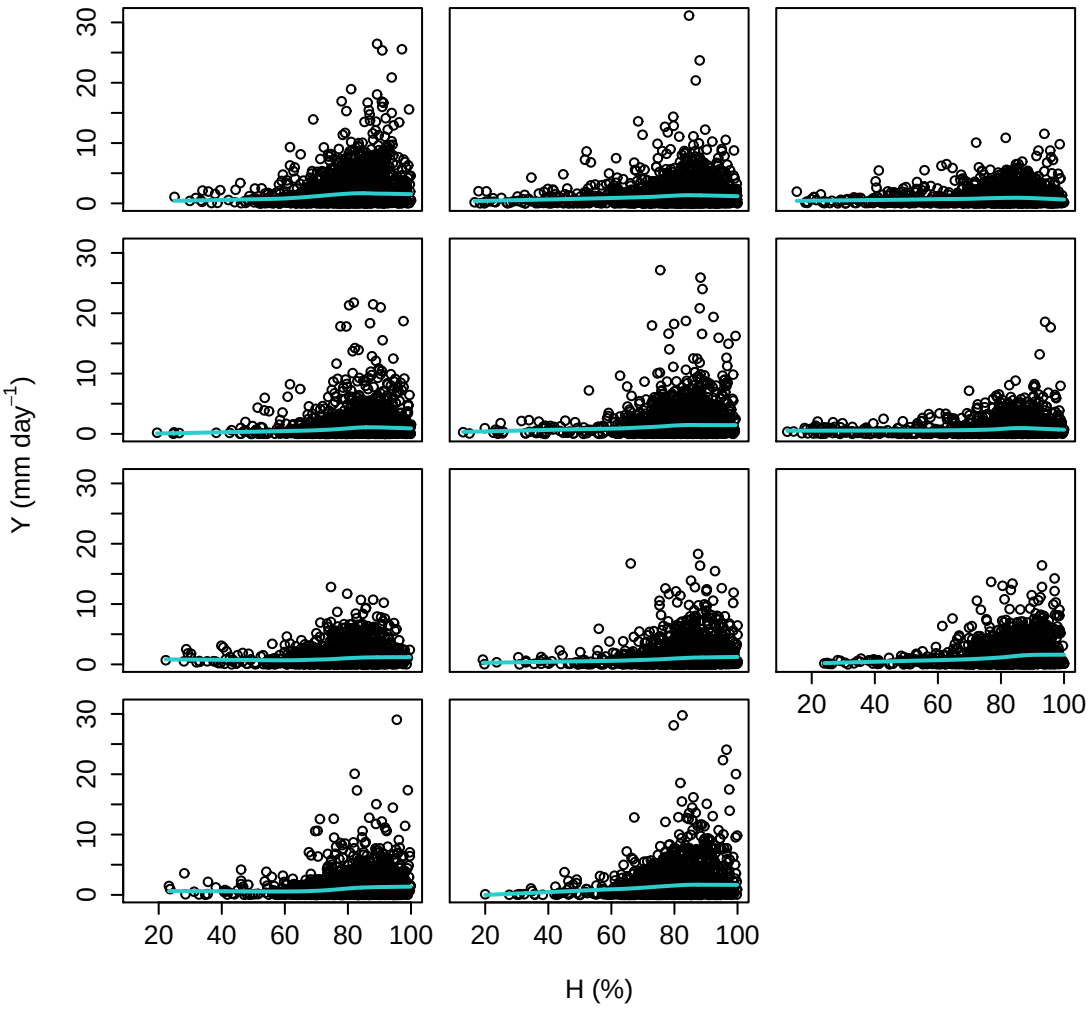

Figure 4.3: Daily-total convective precipitation (excess over 0.1 mm) (Y) against daily-mean relative humidity (H) on wet days in each ensemble member at a single grid point over Birmingham. Curves are local polynomial smoothers. The control run is top left.

uration specific humidity (bottom-left panel) and relative humidity (bottom-right panel), compared to the temporal variation within each ensemble member.

Overall, relative humidity appears to be a positive driver of convective precipitation, as does lapse rate over most values. Saturation specific humidity appears to have little effect on convective precipitation on its own, but Figure 4.4 suggests that this is not the case when considered along with the other predictors. There seems some variation across ensemble members in the relationships between the predictors and the convective precipitation, but perhaps little compared to the temporal variation in the precipitation. The box plots of Figure 4.5 imply that the uncertainty in the parameters of the RCM does not result in substantial variation across ensemble members in the convective precipitation bias. However to quantify the extent to which the simulation of convective precipitation is robust to perturbing the parameters of the RCM in a more general sense, the methodology of Chapter 2

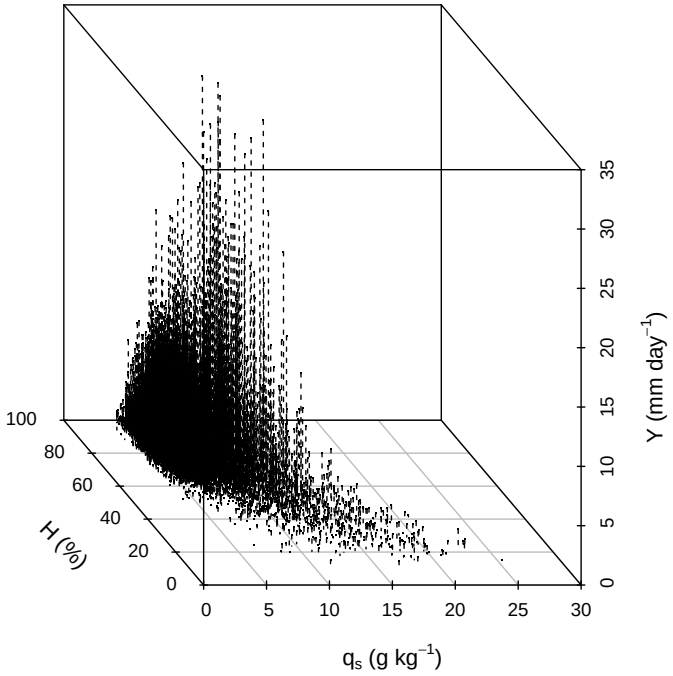

Figure 4.4: Daily-total convective precipitation (excess over 0.1 mm) (Y) against daily-mean saturation specific humidity (q_s) and daily-mean relative humidity (H) on wet days in all ensemble members pooled at a single grid point over Birmingham.

is required.

4.4 Statistical tools

As we outline in the next section, we use spline functions to capture the non-linear relationships between the predictors and precipitation seen in the exploratory data analysis of the previous section. We also use quantile residuals (Dunn and Smyth, 1996) to assess the validity of the assumptions that underpin the statistical model at the Birmingham grid-point. In this section we provide some background on spline functions and on quantile residuals.

4.4.1 Spline functions

A spline function of a predictor is a smooth function that allows a non-linear relationship between the predictor and the response to be captured in a statistical model. Splines are piecewise regression functions that are constrained to join at points called knots to form spline functions. These knots could be placed at specified quantiles of a predictor, for example. The simplest regression splines are linear functions joined at the knots, but spline functions are generally more complex. Certain decisions have to be made for fitting spline smoothers. These decisions include

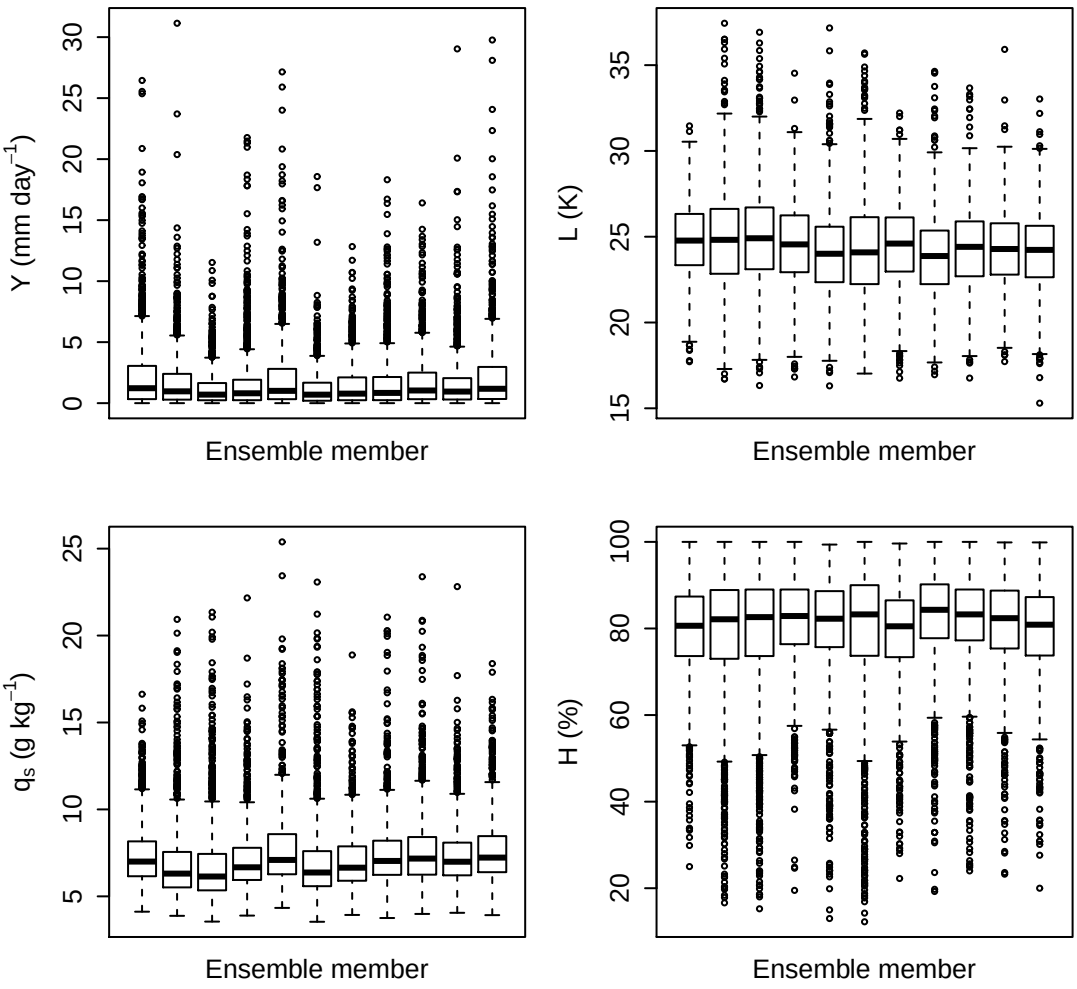

Figure 4.5: Box plots of daily-total convective precipitation (excess over 0.1 mm) (Y) (top-left panel), daily-mean lapse rate (L) (top-right panel), daily-mean saturation specific humidity (q_s) (bottom-left panel) and daily-mean relative humidity (H) (bottom-left panel) at a single grid point over Birmingham, by ensemble member. The control run is on the left of each panel.

the order of the polynomials of the piecewise functions, the number of knots, and the location of the knots. For some types of splines the order of the polynomials is set, and the number of knots control the flexibility of the function. The fewer the number of knots used, the smoother (less local curvature) the function will be because every knot allows more complexity. However some types of splines also have a smoothing parameter that controls flexibility (see Wand, 2000). Overall, there are many different types of splines, including regression splines, cubic splines, B-splines, P-splines, natural splines, thin-plate splines, and smoothing splines, as well as various different combinations. Each type of spline has advantages and disadvantages, and so the type of splines that is used should depend on the application. We focus on cubic splines and natural splines below.

In general, the model matrix of a statistical model is the matrix of its predictors (including a constant for the intercept), such that each column contains the values of

a predictor when linear relationships are present. The basis of a predictor function used in a statistical model can be thought of as the columns of the model matrix that are required to define the function. For example a linear relationship would have a basis of one column, whereas a quadratic relationship requires a basis of two columns, etc (given that an overall intercept is assumed to already be included in the model matrix). With splines, a basis must be defined for every piecewise function. Specifically, the basis of a spline function must contain the number of columns of the order of the polynomials being used (e.g. three for cubic), and a further column for every extra piecewise function. Since the piecewise functions are separated by knots, the number of piecewise functions of a spline function is defined by the number of knots.

Spline functions of third order are relatively common. This means that cubic polynomials are used for the piecewise functions that connect the knots. For example, suppose that a response variable Y is regressed on a single predictor X, both of which vary with time (indexed by i). Also suppose that the relationship between X and Y is non-linear, such that the non-linearity cannot be captured by a single cubic functional form, but can be captured by placing two knots (c_1 and c_2) at specified quantiles of X and using piecewise cubic functions of X between and outside the knots. In this example, a linear regression model of the form

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \beta_4 (X_i - c_1)_+^3 + \beta_5 (X_i - c_2)_+^3 + \epsilon_i$$

could be specified, where $\beta_0, ..., \beta_5$ are fixed parameters and ϵ is the residual term; the $+$ subscripts indicate that the term is set to zero if otherwise negative. The regression model above highlights that the relationship between X and Y is modelled as a cubic polynomial, but one that changes its cubic coefficient for each knot that is exceeded by the value of X. The result is that the relationship is more flexible than only a rigid cubic polynomial with no knots because there are three piecewise polynomials, but that the relationship is continuous and smooth because the piecewise polynomials are related, i.e. smoothly joined at the knots. More knots (and so more piecewise polynomials) can be added if more flexibility is required, with an extra parameter required for each knot added.

To specify a spline function, the number of knots and their placement must be decided upon. Theory or exploratory plots can dictate the placement of knots in any given application, but knots are often simply spaced out evenly across quantiles of the predictor. Doing this reduces the choice to the number of knots used. Choosing the number of knots can be informed by examining exploratory plots of the relationship, or by trial and error until enough detail is captured for an analysis but the relationship is not judged to be over-fitted. Choosing the number of knots can be

done more automatically, however. Since the addition of a knot requires an extra parameter, there is a trade-off between closeness of fit and complexity, which can be judged using information criteria (see Sakamoto et al., 1986) such as optimizing the Akaike information criterion (AIC).

As alluded to previously, many different adjustments can be made to the cubic splines described above. These include natural splines (see Hastie, 1992), which add boundary knots beyond which the function is constrained to linearity. These boundary knots are usually placed at (or close to) the empirical range of the predictor, and are used to allow the relationship to be extrapolated beyond the range of the empirical data more reliably. This is because otherwise the piecewise polynomials outside the outer knots are not necessarily designed to fit well outside the empirical range of the predictor. Although assuming linearity outside the empirical range is not guaranteed to fit well either, it reduces the impact of the outer polynomials, which could be strongly non-linear and are only designed to connect the knots, not to be extrapolated. See (e.g.) Green and Silverman (1994) and Wood (2006) for further detail on splines and alternative methods.

4.4.2 Quantile residuals

There are many different types of residuals that can be used to assess the validity of the assumptions that underpin statistical models. The simplest residuals are raw residuals (empirical data minus the fitted value of each). However raw residuals, as with many other types of residuals, are less informative when the shape and variance of the fitted distributions of a GLMM vary, which is the case with (e.g.) Gamma GLMMs. This is because residuals are often examined by comparing them to a Gaussian distribution, and residuals such as raw residuals will not necessarily follow a Gaussian distribution even if the assumptions that underpin the GLMM are valid. In contrast, quantile residuals (Dunn and Smyth, 1996) are Gaussian if the GLMM assumptions are true for any GLMM (albeit that some randomization is required when the response variable is discrete, e.g. Poisson).

Quantile residuals are computed by inverting the fitted distribution function for each value of the response variable, and then computing the equivalent quantile of the standard Gaussian distribution (i.e. the Gaussian distribution with mean zero and variance one). Let μ and λ be the parameters of the fitted distribution for empirical response y, and let $F(y_i; \mu_i, \lambda_i)$ be the fitted cumulative distribution function for the ith value of y. The ith quantile residual is

$$r_i = \Phi^{-1}[F(y_i; \mu_i, \lambda_i)] \, ,$$

where Φ is the cumulative distribution function of the standard Gaussian. If the statistical model is well specified, then $F(y; \mu, \lambda)$ will come from a uniform distribution on the unit interval, and the quantile residuals will come from a standard Gaussian distribution.

4.5 Statistical model specification

Following the methodology of Chapter 2, we require a statistical model at each grid point to quantify the sources of variation. A Binomial and Gamma hierarchical statistical model (see Coe and Stern, 1982) based on generalized linear models (GLMs) (Nelder and Wedderburn, 1972) is often used to analyse daily or pentad (five-day mean) precipitation, i.e. the probability of wet-day occurrence and the level of precipitation on wet-days (e.g. Buishand and Klein Tank, 1996; Sapiano et al., 2006). In this study we are focusing on wet-day convective precipitation totals, and so we use the Gamma part of this hierarchical statistical model. More specifically, we are analysing wet-day convective precipitation totals in an ensemble from an RCM (such that we have a time series from each member), and so we use a Gamma GLMM to allow the relationships between the predictors and the precipitation to vary with ensemble member via random effect terms ($\boldsymbol{\psi}$). The Gamma family of distributions is used to capture the positive skew of daily-total convective precipitation seen in the exploratory data analysis over the Birmingham grid-point (Section 4.3).

As seen in Section 4.3, the relationships between the predictors and convective precipitation can be non-linear at the Birmingham grid-point. Furthermore we use the same statistical model specification (fitted separately) at each grid point over southern Britain, and so the specification must be flexible enough to capture the relationships at any of these grid points. As a result, we use natural cubic spline bases (described in Subsection 4.4.1) to capture the relationships between the predictors and the precipitation, i.e. we use smooth functions of the predictors to capture the relationships. We use seven parameters for each of these smooth functions at each grid point, which allows us to capture non-linear relationships and provides greater flexibility than assuming linearity or a more rigid functional form such as cubic relationships. Let f_L, f_{q_s} and f_H denote these natural spline functions that relate lapse rate, saturation specific humidity and relative humidity respectively to the linear predictor of the GLMM. We use natural splines because the decomposition requires the expected (μ) daily-total convective precipitation to be estimated outside the empirical range of the data, and natural splines are designed for this by assuming a linear relationship beyond specified outer knots.

It follows from the theory set out in Section 4.1, and the exploratory data analysis

of Section 4.3, that the predictors should be allowed to interact when driving convective precipitation. This can seen because convective instability without moisture will generally lead to little precipitation, similarly a high amount of moisture but little convective instability will generally lead to little convective precipitation. It requires convective instability and the availability of moisture to drive convective precipitation. We account for this interaction between the predictors by using the log link function to relate them to the expected convective precipitation μ, thereby ensuring a multiplicative effect of the predictors on the precipitation.

The GLMM specification is

$$Y_{ij} \mid \left(\boldsymbol{\psi}_j, L_{ij}, q_{sij}, H_{ij}\right) \sim \text{Gamma}(\mu_{ij}, \lambda_{ij})$$

$$\begin{aligned}\log(\mu_{ij}) &= \beta_0 + \psi_j^{(0)} + f_L(L_{ij}, \psi_j^{(L)}) + f_{q_s}(q_{sij}, \psi_j^{(q_s)}) + f_H(H_{ij}, \psi_j^{(H)}) \\ \log(\lambda_{ij}) &= \alpha_0 + f_L^{(\lambda)}(L_{ij}) + f_{q_s}^{(\lambda)}(q_{sij}) + f_H^{(\lambda)}(H_{ij})\end{aligned}$$

$$\boldsymbol{\psi}_j = \left(\psi_j^{(0)}, \psi_j^{(L)}, \psi_j^{(q_s)}, \psi_j^{(H)}\right) \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}) \,,$$

where i indexes the wet-day within each ensemble member and j indexes the member. As with the smooth functions (f_L, f_{q_s} and f_H) relating the predictors to the linear predictor ($\log(\mu)$) of the GLMM, smooth functions ($f_L^{(\lambda)}$, $f_{q_s}^{(\lambda)}$ and $f_H^{(\lambda)}$) are used to allow the dispersion parameter (λ) of the GLMM to vary with the predictors (also through a log link function to ensure that the interaction between the predictors is more fully captured by the GLMM). This gives further flexibility to each fitted Gamma distribution, although we do not feel the need to allow λ to also vary with random effects.

As can be seen from the specification of the GLMM, the random effects ($\boldsymbol{\psi}$) are assumed to follow a multivariate Gaussian distribution. The random effects can be partitioned into that ($\psi^{(0)}$) that allows the intercept to vary with ensemble member, those ($\psi^{(L)}$) that allow the relationship between lapse rate and convective precipitation to vary with ensemble member, those ($\psi^{(q_s)}$) that allow the relationship between saturation specific humidity and convective precipitation to vary with ensemble member, and those ($\psi^{(H)}$) that allow the relationship between relative humidity and convective precipitation to vary with ensemble member. Since each smooth function includes seven fixed parameters, $\psi^{(L)}$, $\psi^{(q_s)}$ and $\psi^{(H)}$ consist of seven random effects each.

For each smooth function to include seven fixed parameters, six knots are specified for each predictor. These knots determine at which values of each predictor detail is added to capture its relationship with convective precipitation at a given grid point. At each grid point, we typically place four knots equally spaced across the quantiles of each predictor (at the 20th, 40th, 60th and 80th percentiles of each predictor) and one in each tail (e.g. at the 2nd and 98th percentiles) to capture the effect of the extreme values in each predictor. The fit of the GLMM can be sensitive to the placing of these knots in the tails of the distribution of each predictor, and so we base this choice on the placings that lead to the total variation ($\text{var}(Y)$) in the wet-day convective precipitation being estimated relatively closely to its empirical variation ($\text{var}(y)$). These knot placings ensure that detail is captured in the relationship between each predictor and convective precipitation across the quantiles of each predictor, and that the empirical variation in the convective precipitation is captured relatively well by the GLMM.

We estimate the parameters of each GLMM by the method of interconnected GLMs of Lee et al. (2006), which is also summarized and demonstrated by Rönnegård et al. (2010). This method involves using three different GLMs to estimate (by iteratively reweighted least squares) the parameters (β_0 and the fixed parameters in f_L, f_{q_s} and f_H) of the mean part of the model, the parameters (α_0 and the parameters in $f_L^{(\lambda)}$, $f_{q_s}^{(\lambda)}$ and $f_H^{(\lambda)}$) of the dispersion part of the model, and the variance components ($\boldsymbol{\Sigma}$) of the random effects. Each GLM uses output from the others, and so, after initializing the parameters values, can be updated iteratively until convergence. See Lee et al. (2006) and Rönnegård et al. (2010) for details on the fitting algorithm.

With respect to the validity of the assumptions behind each GLMM, Figure 4.6 shows diagnostic plots for the quality of fit of the GLMM at the Birmingham grid-point using quantile residuals (Subsection 4.4.2). The top-left panel of the figure shows that the greatest (positive) quantile residuals are at low fitted values. This suggests that when the GLMM predicts little convective precipitation, it is often an underestimation. There is also an indication that at fitted values between 2 and 5 mm day^{-1}, the GLMM can greatly overestimate the amount of convective precipitation. Despite this, the variation in the quantile residuals seems constant across the distribution of fitted values (see top-right panel), there seems no temporal variation in the quantile residuals (centre-left panel), and no variation across ensemble members in the distribution of the quantile residuals (centre-right panel). The residuals appears to approximate a standard Gaussian distribution (bottom panels of Figure 4.6) except for in the negative tail of the quantile residuals, which is too heavy. This suggests that the greatest (by fitted probability) overestimates are too great, i.e. that certain atmospheric conditions lead to the convective precipitation being

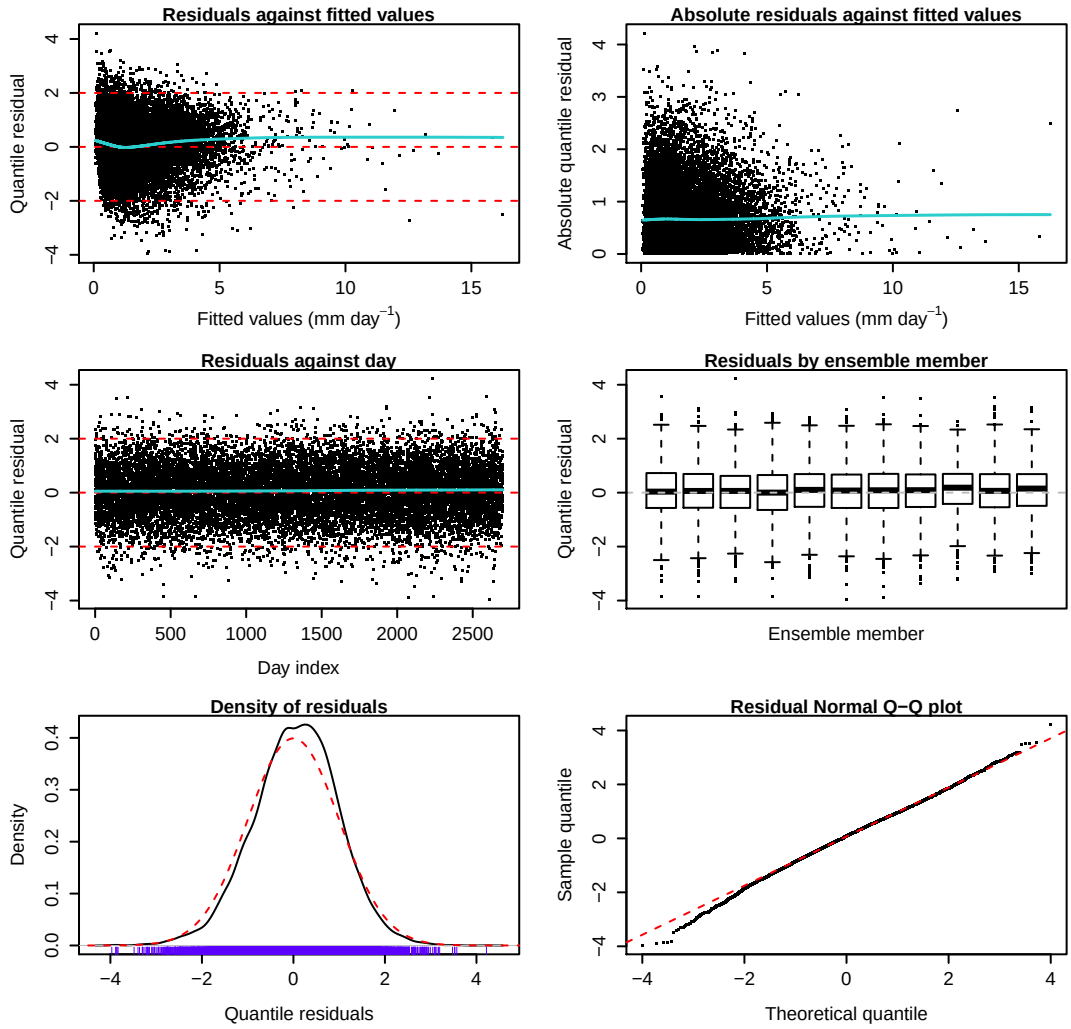

Figure 4.6: Diagnostic residual plots (using quantile residuals) for the GLMM at a single grid point over Birmingham. The dashed red lines provide an idealized guide, and the solid blue curves are local polynomial smoothers showing any trends where appropriate.

(systematically) overestimated.

Further diagnosing the quality of the GLMM fit at the Birmingham grid-point, Figure 4.7 compares plots of the empirical (top panels) convective precipitation against the three predictors, and similar plots for values (bottom panels) of convective precipitation simulated from the GLMM. The relationships (including their variation across the ensemble members) and general shape of the distribution of convective precipitation given each predictor appear to be captured well by the GLMM. However it does seem that the Gamma family can be too heavy tailed, such that the GLMM can simulate days of convective precipitation that are more extreme than seen in the empirical data. Despite this, Figures 4.6 and 4.7 indicate that overall the GLMM fits the empirical well enough to use it to decompose the total variation in the wet-day convective precipitation, and is therefore fit for purpose.

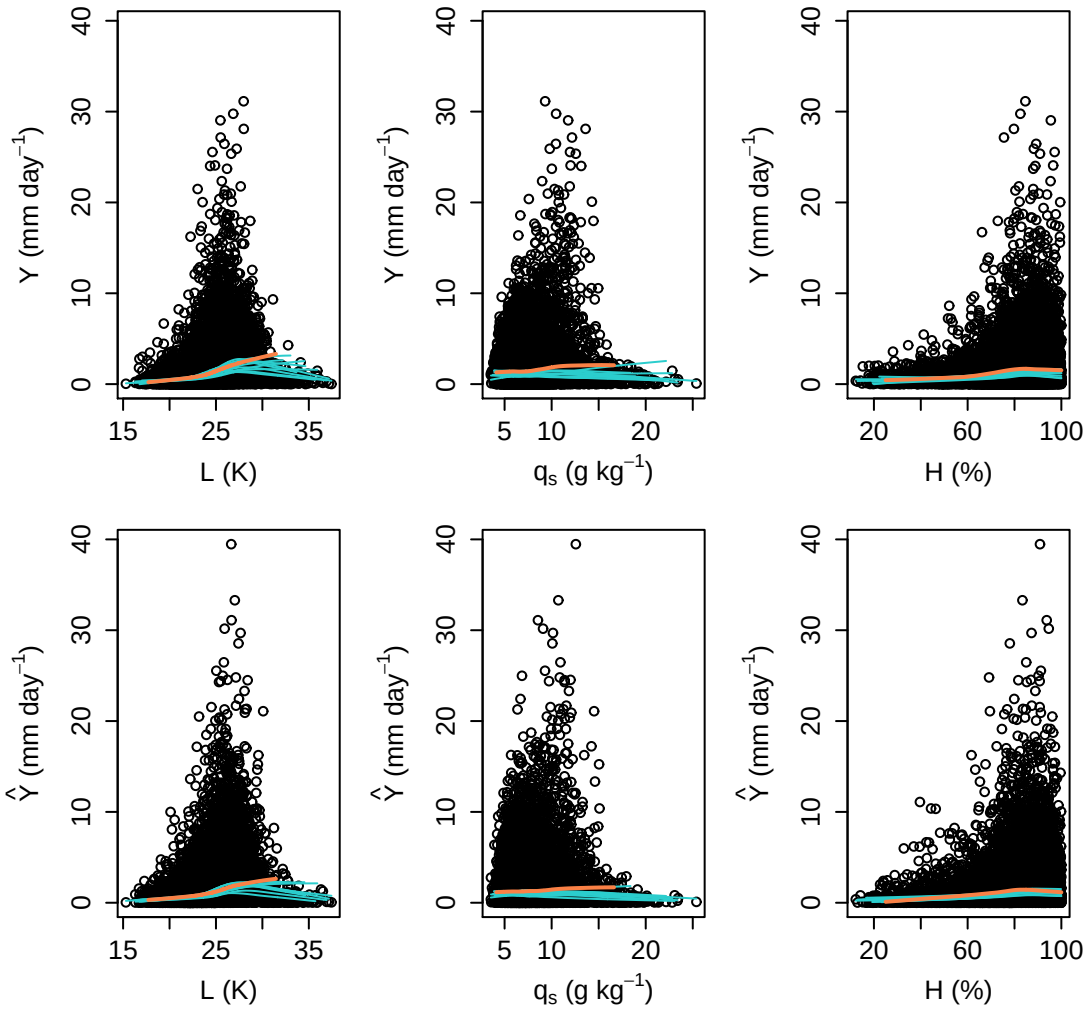

Figure 4.7: GLMM diagnostics: Top panels are the response (Y) of the GLMM against the predictors (L, q_s and H) at a single grid point over Birmingham. Bottom panels show simulated values ($\hat{Y}$) for Y from the GLMM. Curves are local polynomial smoothers for each ensemble member, with the control run in orange.

4.6 Decomposition

We continue to follow the methodology of Chapter 2 by separating the predictors ($\boldsymbol{X}$) into their means ($\boldsymbol{\gamma}$) within each ensemble member and their temporal deviations ($\boldsymbol{X'}$) from the means, i.e. (including subscripts)

$$\begin{aligned} L_{ij} &= \gamma_j^{(L)} + L'_{ij} \\ q_{sij} &= \gamma_j^{(q_s)} + q'_{sij} \\ H_{ij} &= \gamma_j^{(H)} + H'_{ij} \, . \end{aligned}$$

As such, $\boldsymbol{\gamma}$ denotes the vector of $\gamma^{(L)}$, $\gamma^{(q_s)}$ and $\gamma^{(H)}$, and $\boldsymbol{X'}$ denotes the vector of L', q'_s and H'.

Given the separation of the predictors into their means within each ensemble member and their temporal components, the inputs to each GLMM are

- the variation across ensemble members in the intercept, captured by $\psi^{(0)}$,
- the variation across ensemble members in the relationship between lapse rate and convective precipitation, $\psi^{(L)}$,
- the variation across ensemble members in the relationship between saturation specific humidity and convective precipitation, $\psi^{(q_s)}$,
- the variation across ensemble members in the relationship between relative humidity and convective precipitation, $\psi^{(H)}$,
- the variation across ensemble members in the distribution of lapse rate, $\gamma^{(L)}$,
- the variation across ensemble members in the distribution of saturation specific humidity, $\gamma^{(q_s)}$,
- the variation across ensemble members in the distribution of relative humidity, $\gamma^{(H)}$,
- the temporal variation in lapse rate, L',
- the temporal variation in saturation specific humidity, q'_s,
- the temporal variation in relative humidity, H',

in addition to the residual term, ϵ.

We estimate the importance measures of individual inputs to gain insight into why the representation of the physical process captured by the GLMM varies across the within the ensemble members. We also group together the random effects $\boldsymbol{\psi}$, the predictor means $\boldsymbol{\gamma}$ and the temporal components $\boldsymbol{X'}$ of the predictors and follow the decomposition summarized in Figure 2.6 (page 65) of Subsection 2.2.2. We do this separately at each grid point. In doing so we quantify the extent to which the simulated convective precipitation is robust to perturbing the RCM parameters in a more general sense than only accounting for bias in the mean rate of the convective precipitation. To emphasize this, we compare the results to those of a random-effects analysis of variance (random-effects ANOVA) (we discuss this in the next section).

Continuing with the description of the decomposition using GLMMs, in the previous section it was outlined that the dispersion parameter λ is allowed to vary with the predictors in the GLMM in this application. Although this does not affect the sampling of the expectation μ, it does affect the simulation when estimating the

total variation (var(Y)) in the convective precipitation. When simulating var(Y) at a given grid point, λ (in addition to μ) must be computed for each sample of the inputs. Each sample of λ and μ must then be used when computing the corresponding sample of Y.

With respect to modelling the inputs, we follow the recommended dependence structure of Subsection 2.2.2 and Section 2.4. Therefore any dependence between the inputs that capture variation across the different ensemble members (i.e. the random effects and predictor means) is accounted for, as is the dependence between the inputs that capture the temporal variation in the represented atmospheric conditions (the temporal components of the predictors), but independence is assumed between these two groups. This assumption can be seen as valid because the box plots of the top-right and bottom panels of Figure 4.5 (page 122) suggest that there is little variation across ensemble members in the shape of the distributions of the predictors. There is some indication that (e.g.) the variance of the predictors is relatively small within the control run (on the left of each panel) of the ensemble, but this variation across ensemble members appears small and many of the distributions of each predictor are similar across the other ensemble members.

Following Section 2.4, we use a multivariate Gaussian copula to represent the random effects and predictor means, with empirical marginals for the latter, and a three-dimensional kernel density with empirical marginals to represent the temporal components of the predictors. The three-dimensional kernel density is required because there is non-linear and asymmetrical dependence between these inputs (see Figure 4.8). Note that this dependence could have been reduced by (e.g.) using principle components analysis (e.g. see Chapter 8 of Mardia et al., 1979) on the predictors before fitting each GLMM, however we did not choose to do this because it may have complicated the interpretation of the predictors (interpreting the effect of a predictor is only useful if the predictor is physically meaningful).

Using the model for the inputs described above, along with the fitted GLMM, we use simulation to estimate the importance measures of Chapter 2. Log link functions relate the mean (μ) and dispersion (λ) parameters of the Gamma distributions to the effects of the inputs in the GLMM at a given grid point. However some samples of μ and λ are generated outside the range of the fitted values of the GLMM, and when this occurs the non-linear nature of the log function can result in unrealistic values of convective precipitation. This is because the log function can be strongly non-linear and so will not necessarily represent the relationship between the inputs and μ well when extrapolated out to greater values than the greatest fitted value of μ (similarly for λ). To avoid this issue, we use natural interpolation splines (see

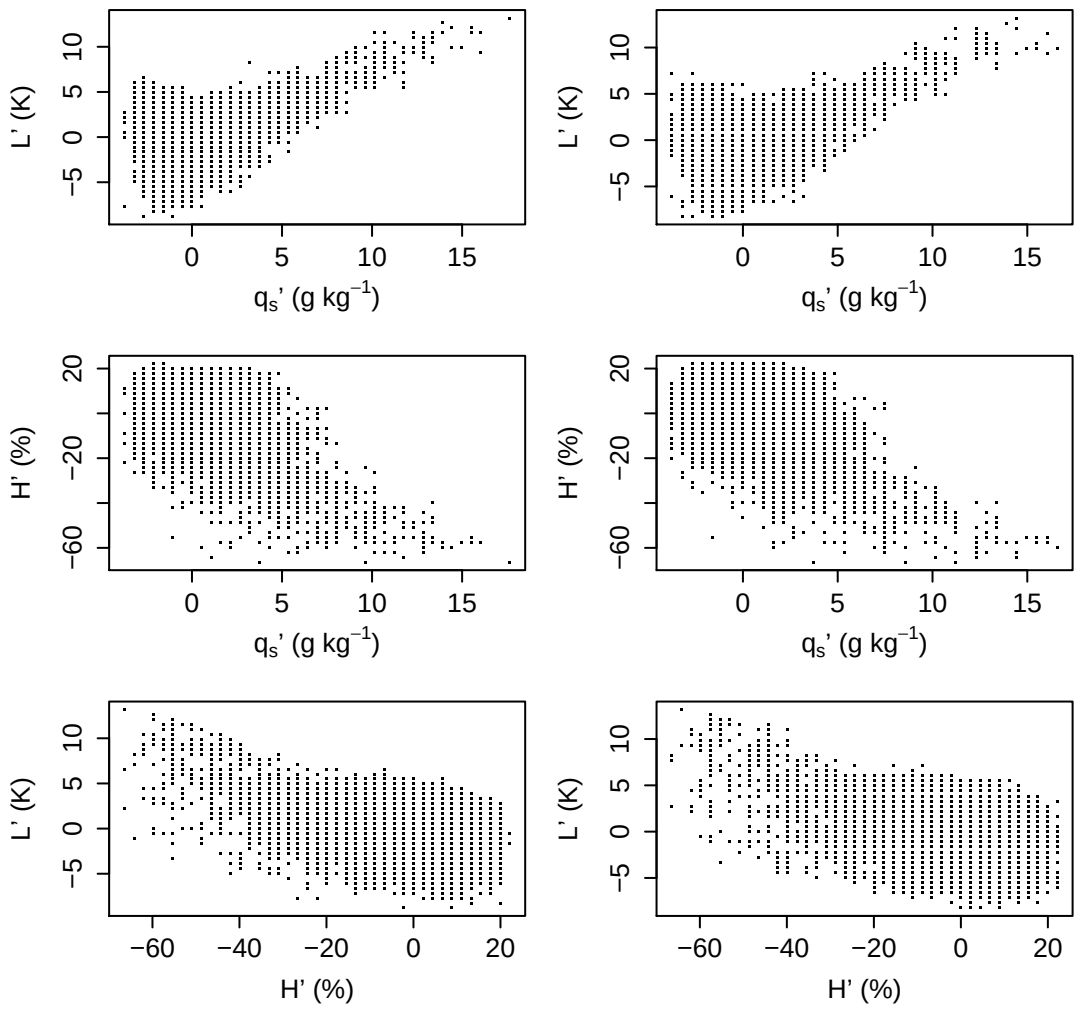

Figure 4.8: Two-dimensional representation of the three-dimensional kernel density estimate used to model the temporal components (L', q_s' and H') of the predictors at the Birmingham grid-point. The empirical data are shown in the left panels, whereas the right panels show samples simulated from the kernel density model; binning is used in all of the plots to allow the shapes of the distributions to be seen more clearly.

Dougherty et al., 1989) when extrapolating to greater values than the greatest fitted value of μ and of λ, instead of the log function. Natural interpolation splines are similar to the natural splines described in Subsection 4.4.1 except that they join all data points, instead of smoothing through random data. As such, we use the natural interpolation splines to assume linearity beyond the range of the empirical fitted values of μ and λ, while retaining a continuous relationship. This is shown by the solid black curve in Figure 4.9 for μ. This linear extrapolation is only relevant to a small proportion of samples, but ensures that the simulated samples for μ and λ are more reliable.

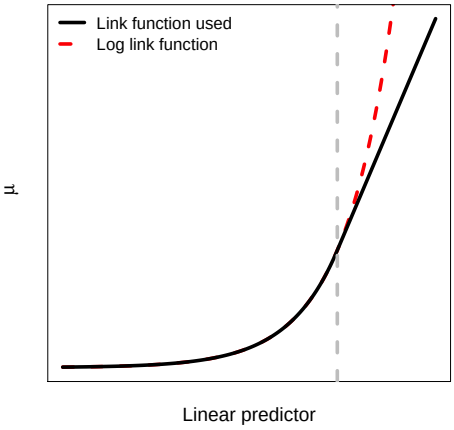

Figure 4.9: Schematic showing the link function (solid black curve) used to relate the effects of the inputs (i.e. the linear predictor) to the fitted expectation (μ) of convective precipitation at any given grid point. The link function used is the log link function (dashed red curve) up to the greatest fitted value (marked by the vertical dashed grey line) at the grid point, but linear at greater values.

4.7 Random-effects ANOVA

As outlined in Sections 1.7 and 2.5, a random-effects ANOVA can be used to estimate the variation across ensemble members in the mean rate of convective precipitation. Instead of the statistical model specification outlined in Sections 1.7 and 2.5, however, perhaps the most comparable ANOVA model to the GLMM outlined in Section 4.5 is the random-effects ANOVA that uses the same family of distributions (i.e. the Gamma family) as the GLMM, and the same fitting algorithm (i.e. that of Lee et al., 2006), but without the predictors. Therefore, to compare to the GLMM, we can use the random-effects ANOVA model

$$\begin{aligned} Y_{ij} \mid \psi_j^* &\sim \text{Gamma}(\mu_j^*, \lambda^*) \\ \log(\mu_j^*) &= \beta_0^* + \psi_j^* \\ \psi_j^* &\sim N(0, \tau_0) \ , \end{aligned}$$

where i indexes time, j indexes ensemble member, μ^* and λ^* are the fitted expectation and dispersion parameters respectively, β_0^* is a fixed parameter, and τ_0 is the variance component of random effect ψ^*. As with the specification outlined in Sections 1.7 and 2.5, the model outlined here is a random-effects ANOVA because the only input (other than the residual term) is a single random effect term that allows for variation across the ensemble members in the fitted expectation. As with the GLMM of Section 4.5, simulation can be used to estimate the result of the random-effects ANOVA, i.e. the total proportion of variation in the precipitation accounted for by the input ψ^*.

Using the random-effects ANOVA outlined in this section at the Birmingham grid-point, for example, 2% of the total variation in the convective precipitation is estimated to be due to variation across ensemble members in mean precipitation rate. As a part of the results reported in the next section, we use the random-effects ANOVA and compare its result to the variation that we find to be due to the uncertainty in the RCM parameters using GLMMs. As discussed in Section 2.5, unlike the random-effects ANOVA, the GLMMs use predictors to capture temporal variation in the precipitation. As a result, we expect to identify more variation to be due to the uncertainty in the RCM parameters using GLMMs because the uncertainty will not necessarily only affect the mean rate of convective precipitation. This allows us to more fully identify and examine the impact of uncertainty in RCM parameters on the simulation of convective precipitation.

4.8 Results

4.8.1 Single grid point

Before analysing maps of southern Britain, we demonstrate the results at the Birmingham grid-point. Figure 4.10 shows the decomposition of the total variation in daily-total convective precipitation in the perturbed physics ensemble at the grid point. The decomposition is based on 1500 replications of samples of size 2000 (i.e. $R = 1500$ and $r = 2000$ using the notation of Section 2.3). The total variation in the precipitation is estimated higher than the sample variance by a ratio of 1.22 at this grid point. The Gamma family was found to be (slightly) too heavy tailed in Section 4.5, and it appears that this has contributed to the total variation in convective precipitation being overestimated. Despite this, the relationships between the predictors and the convective precipitation were found to represent that of the empirical data well by Figure 4.7, which suggests that the heavy tailed Gamma fit may result in the residual variation being overestimated instead of the other importance measures.

Of the total variation in the convective precipitation, an estimated 33% is accounted for by the GLMM. The predictors and their relationships with convective precipitation are designed to capture the representation by the RCM of the fundamental driving physical process of convective precipitation on a local spatial and temporal scale (i.e. that convective instability of moist air leads to convective precipitation). However the predictors (and additional precipitation bias) are only able to capture 33% of this variation. This could be because convective instability or moisture may be at different atmospheric heights to those represented, or because the tempo-

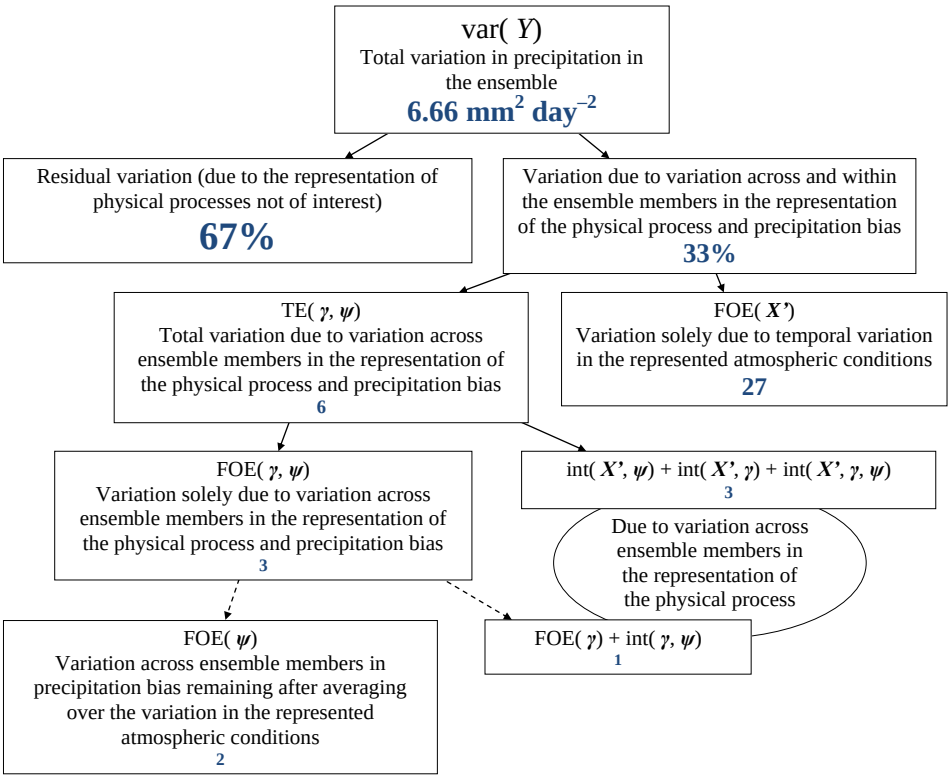

Figure 4.10: Decomposition of daily-total convective precipitation (Y) for the Birmingham grid-point. Each total effect (TE) (on μ), first-order effect (FOE) and interaction (int) is expressed as a percentage of var(Y) (rounded to integers). Dashed lines indicate where non-orthogonality could have occurred due to dependence between inputs. The empirical variance of Y is 5.46 mm^2 day^{-2}.

rally localized nature of convective precipitation results in detail being lost by the temporal resolution of the data (i.e. the relationships between the predictors and precipitation may be strongest on the sub-daily scale).

The majority of the variation accounted for by the GLMM is due to the temporal variation in the represented atmospheric conditions. Specifically, of the total variation in the convective precipitation, 27% is due to the temporal variation (FOE($\boldsymbol{X'}$)) in the predictors. With respect to the differences between ensemble members, a total (TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$)) of 6% is estimated to be due to the uncertainty in RCM parameters. Given the predictors selected, the aim was to fully quantify the extent to which the simulation of convective precipitation events is robust to the uncertainty in the RCM parameters on a local spatial and temporal scale. However the implication of only accounting for 33% of the total variation in the convective precipitation with the GLMM is that this variation found to be due to the uncertainty in RCM parameters is likely to be underestimated. This is because it is unknown how much of the remaining (unaccounted for) 67% of the variation is due to the uncertainty in the RCM parameters and how much is temporal (internal) variation.

The 6% found to be due to the uncertainty in RCM parameters may not seem

large, but using a random-effects ANOVA (described in the previous section) only identifies 2% of the total variation in the convective precipitation to be due to this uncertainty in the RCM parameters. This is because the largest impact of perturbing the RCM parameters is on the relationship between the represented atmospheric conditions and convective precipitation, and much of this variation is not seen in the mean precipitation rate. To see this, first consider the first-order effect (FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$)) of the predictor means and random effects of the GLMM, and the first-order effect (FOE($\boldsymbol{\psi}$)) of the random effects. As outlined in Section 2.5, FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) should recover the result from a random-effects ANOVA. Although Figure 4.10 shows FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) estimated to be 3%, it is rounded to the zero decimal places from 2.51%, whereas the ANOVA result is rounded down from 2.25%, so the discrepancy is small (note that TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) is rounded from 5.90%).

At this grid point, FOE($\boldsymbol{\psi}$) recovers the result of a random-effects ANOVA. This is because there is little variation across ensemble members in the frequency distribution of the atmospheric conditions represented by the predictors (FOE($\boldsymbol{\gamma}$) + int($\boldsymbol{\gamma}, \boldsymbol{\psi}$) = 1%). Given that FOE($\boldsymbol{\gamma}, \boldsymbol{\psi}$) is estimated to be 2.51%, it is clear that much of the variation found to be due to the uncertainty in the RCM parameters is driven by variation in the relationship between the represented atmospheric conditions and convective precipitation. Overall, by using predictors with a GLMM we have found substantially more disagreement (6% of the variation) in the simulation of convective precipitation that is due to uncertainty in RCM parameters than would be identified (2%) by using a random-effects ANOVA.

To provide further insight into why the physical process depends on the RCM parameters, and insight into the dominant temporal components of the physical process, we display the importance measures of the individual inputs in Table 4.2. Dependence between the inputs means that the importance measures of the individual inputs are non-orthogonal to each other, but the top three rows of the table suggest that the variation in lapse rate contributes more to the temporal variation in the driving physical process of convective precipitation than the variation in saturation specific humidity or relative humidity. The total effect (21%) and first-order effect (12%) of the temporal component in lapse rate are greater than those of the humidity variables. Indeed, little of the total variation in daily-total convective precipitation appears to be solely due to the temporal variation in either saturation specific humidity (first-order effect 2%) or that in relative humidity (2%). This is perhaps not surprising because, as noted in Section 4.3, these humidities need to be considered in conjunction with each other, and lapse rate, to capture the physical process. Indeed, the total effects of the temporal components of saturation specific humidity (8%) and relative humidity (12%) suggest that both humidity variables

are important to the convective precipitation when considered in conjunction with the other inputs. Conversely, the importance of the variation in lapse rate is not so dependent on the other predictors, which indicates that there is usually enough moisture in the air for a high lapse rate to result in convective precipitation.

Input	TE (%)	FOE (%)
L'	21	12
q_s'	8	2
H'	12	2
$\gamma^{(L)}$	0	0
$\gamma^{(q_s)}$	0	0
$\gamma^{(H)}$	0	0
$\psi^{(0)}$	1	2
$\psi^{(L)}$	1	1
$\psi^{(q_s)}$	1	1
$\psi^{(H)}$	0	1

Table 4.2: Total effect (TE) (on expectation μ) and first-order effect (FOE) of individual inputs when decomposing the total variation in daily-total convective precipitation (Y) at the Birmingham grid-point. Values are expressed as percentages of var(Y) and rounded to zero decimal places.

The fourth to sixth rows of Table 4.2 show that the simulation of all three predictors is robust to the uncertainty in the RCM parameters. Each rounded to the nearest percent, no variation in convective precipitation is found to be due to the variation across ensemble members in the distribution of any of the predictors. This shows that the frequency distribution of the represented atmospheric conditions is robust to the uncertainty in the RCM parameters.

With respect to the random effects, for the reasons outlined in Section 2.1 we do not advocate using the first-order effects of the individual random effects in this application, other than to estimate their interactions with the other inputs (by comparing the first-order effect of a random effect to its total effect). Focusing on the bottom four rows of Table 4.2, the first-order effects can be greater than the total effects. This highlights that the dependence between the inputs inflates the first-order effects of the random effects relative to their total effects. Nethertheless, these importance measures suggest that the variation in convective precipitation found to be due to the uncertainty in the RCM parameters is not dominated by the variation across ensemble members in any single relationship between a predictor and the precipitation, and that the variation in each of the relationships could contribute.

Overall, by attempting to capture the representation by the RCM of the fundamental

driving physical process of convective precipitation over Birmingham with a GLMM, we find that more variation in daily-total convective precipitation is due to the uncertainty in the parameters of the RCM than would be detected by focusing on only the mean rate of convective precipitation in each ensemble member. More specifically, we find 6% of the total variation in the convective precipitation to be due to the uncertainty in the RCM parameters, which is three times the 2% that is found by using a random-effects ANOVA to quantify the variation across ensemble members in precipitation bias. This is largely because of variation across ensemble members in the relationship between the represented atmospheric conditions and the convective precipitation. Furthermore the 6% that we find to be due to the differences between the ensemble members is likely to be an underestimation because much of the variation in the convective precipitation is unaccounted for by our GLMM. Lapse rate, saturation specific humidity and relative humidity are found to be simulated robustly across ensemble members. The temporal variation in lapse rate can be seen as more important to convective precipitation than that in humidity because there is often enough moisture in the air to mean that convective instability is the dominant driver.

4.8.2 Exploratory maps over southern Britain

To examine the spatial pattern of importance measures, we repeat the decomposition at each grid point (separately) over southern Britain. Before focusing on these results, it is worth examining some exploratory maps over the region. Figures 4.11, 4.12 and 4.13 show such maps: Exploratory maps of the convective precipitation are shown in Figure 4.11, exploratory maps of the predictors (along with mean wind velocities, and orography) are shown in Figure 4.12, whereas Figure 4.13 shows correlations between various variables. We examine many of the plots in these figures in this subsection to characterize the study region, and refer to others from the next subsection to provide a physical interpretation of our results.

Figure 4.11 shows that the greatest mean and variance of summer wet-day convective precipitation is along the west coast of southern Britain. Examining the exploratory maps of the predictors shown in Figure 4.12, the high mean convective precipitation along the west coast is despite a relatively weak mean lapse rate (top-left panel). The air in this area can hold relatively little moisture on average (top-centre panel), with the high precipitation rate due to high mean (top-right panel) relative humidity caused by the air ascending up the land from the sea (see bottom-right panel for orography of the land). Mean saturation specific humidity and mean relative humidity follow opposing spatial patterns, with high saturation specific humidity and low relative humidity in the warmer south of the region.

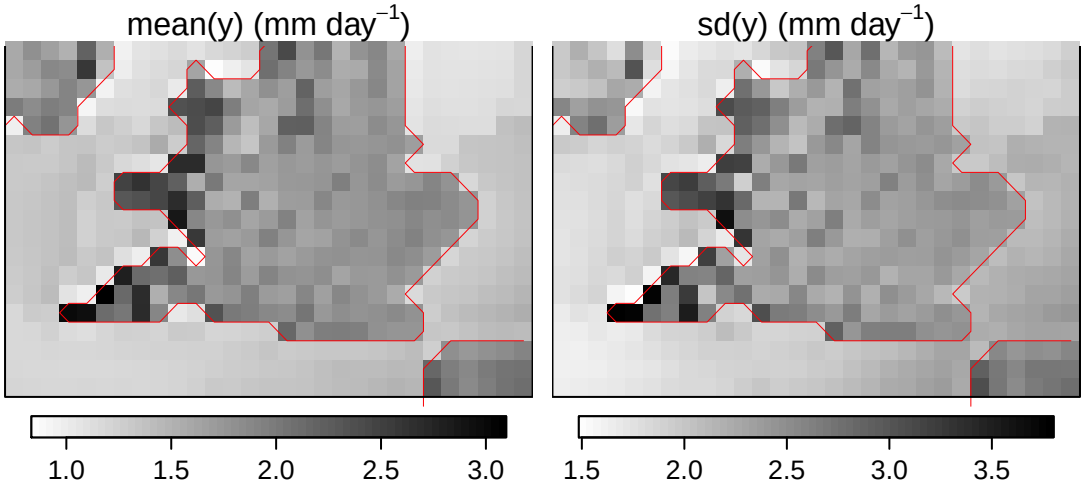

Figure 4.11: Mean (left panel) and standard deviation (sd) (right panel) of the empirical daily-total wet-day convective precipitation (y) at the grid-point level across southern Britain. Coastline is in red.

In addition to the high rates of convective precipitation along the west coast of southern Britain, the top-left panel of Figure 4.11 shows that there is generally more summer convective precipitation over land than over sea. This is because land heats up quicker than sea each day in summer, and this leads to more convection over land (except for at high altitude, where the land is cooler). This can be seen in the mean lapse rate for a given latitude. The greater capacity of the sea to retain heat does, however, mean that the air can on average hold the most moisture over the sea to the south-east, where the air is sheltered somewhat by the land from the cool westerly winds from the sea but warmed by southerly winds from France. This can be seen from the mean wind velocity (bottom-left and bottom-centre panels) over the region and how it correlates (top-left and centre-left panels of Figure 4.13) to temperature at 850 hPa, and also leads to low relative humidity in this area. The centre-left panel of Figure 4.13 also highlights that the correlation between the northward component of wind velocity and temperature is relatively weak immediately off the southern coastline of Britain. This could be due to sea breezes, which are winds caused by the difference between land and sea in their capacity to retain heat, and often result in wind in different directions at different times of each day.

The high variance of all three predictors (centre-left, centre-centre and centre-right panels of Figure 4.12) over the sea to the south-east is a result of relatively high variance (not shown) of the strength of the westerly winds. The variance of lapse rate and relative humidity are also greater near the west and south coasts than further inland and along the east coast due to the variance of the winds and their interaction with the land.

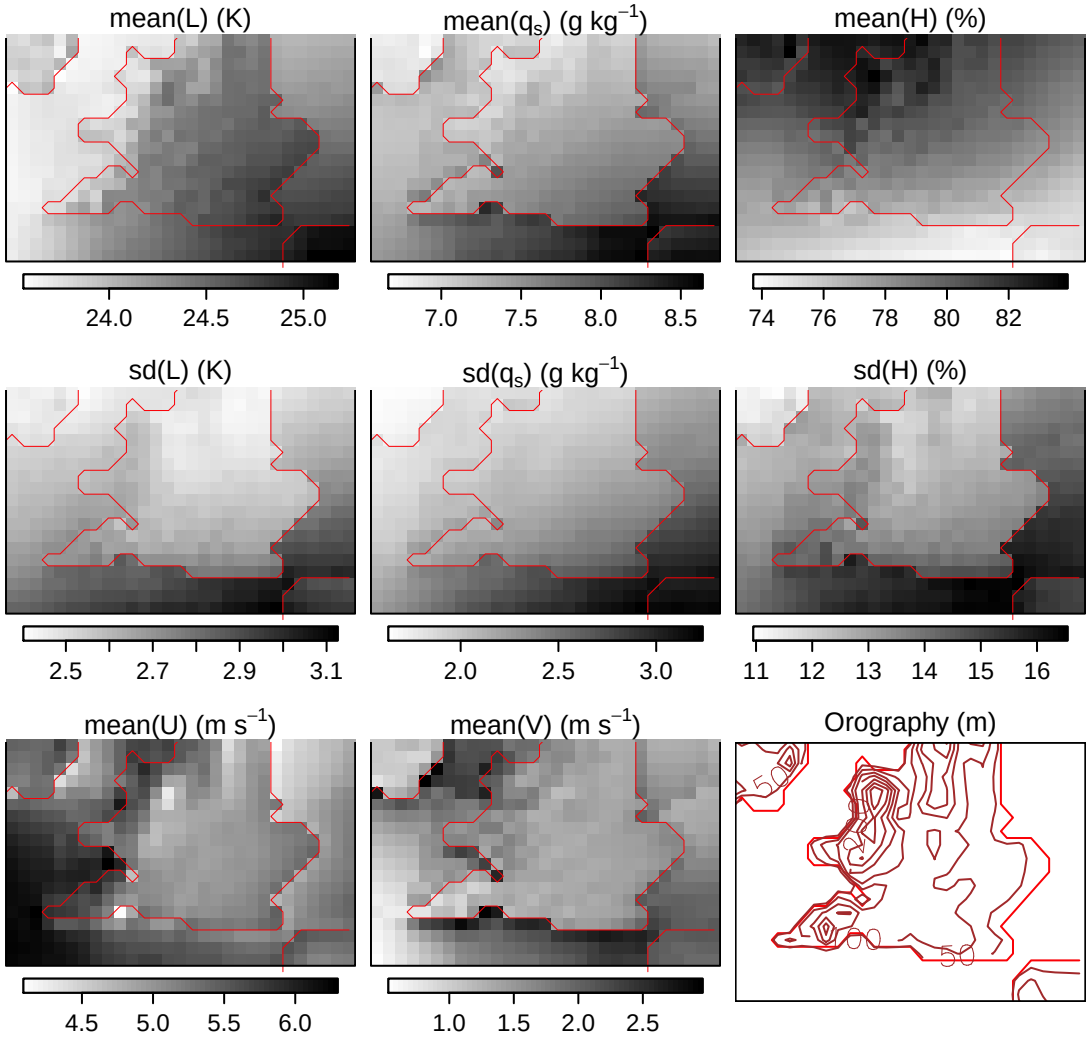

Figure 4.12: Mean of the empirical daily-mean lapse rate (L), saturation specific humidity (q_s), relative humidity (H), eastward component of wind (U) and northward component of wind (V), all at 850 hPa where applicable, on wet days at the grid-point level across southern Britain, in addition to the standard deviation (sd) of L, q_s and H. Orographic contours of the study region are shown in the bottom-right panel. Coastline is in red.

4.8.3 Results maps over southern Britain

Comparing the estimated total variation (var(Y)) in daily-total summer wet-day convective precipitation to that (var(y)) seen in the empirical data, the top panels in Figure 4.14 show that the spatial pattern of the empirical data is captured by the statistical modelling, but that it is generally estimated to be greater than the empirical variation, especially over the land. This was seen over Birmingham (Subsection 4.8.1), likely to be due to the Gamma family being (slightly) too heavy tailed. We describe the spatial patterns of various importance measures below, but it is important to note that the spatial pattern of var(Y)/var(y) is not similar to that of any of the importance measures. We use this as a diagnostic statistic because it highlights that the results we report below are not simply a result of the GLMMs

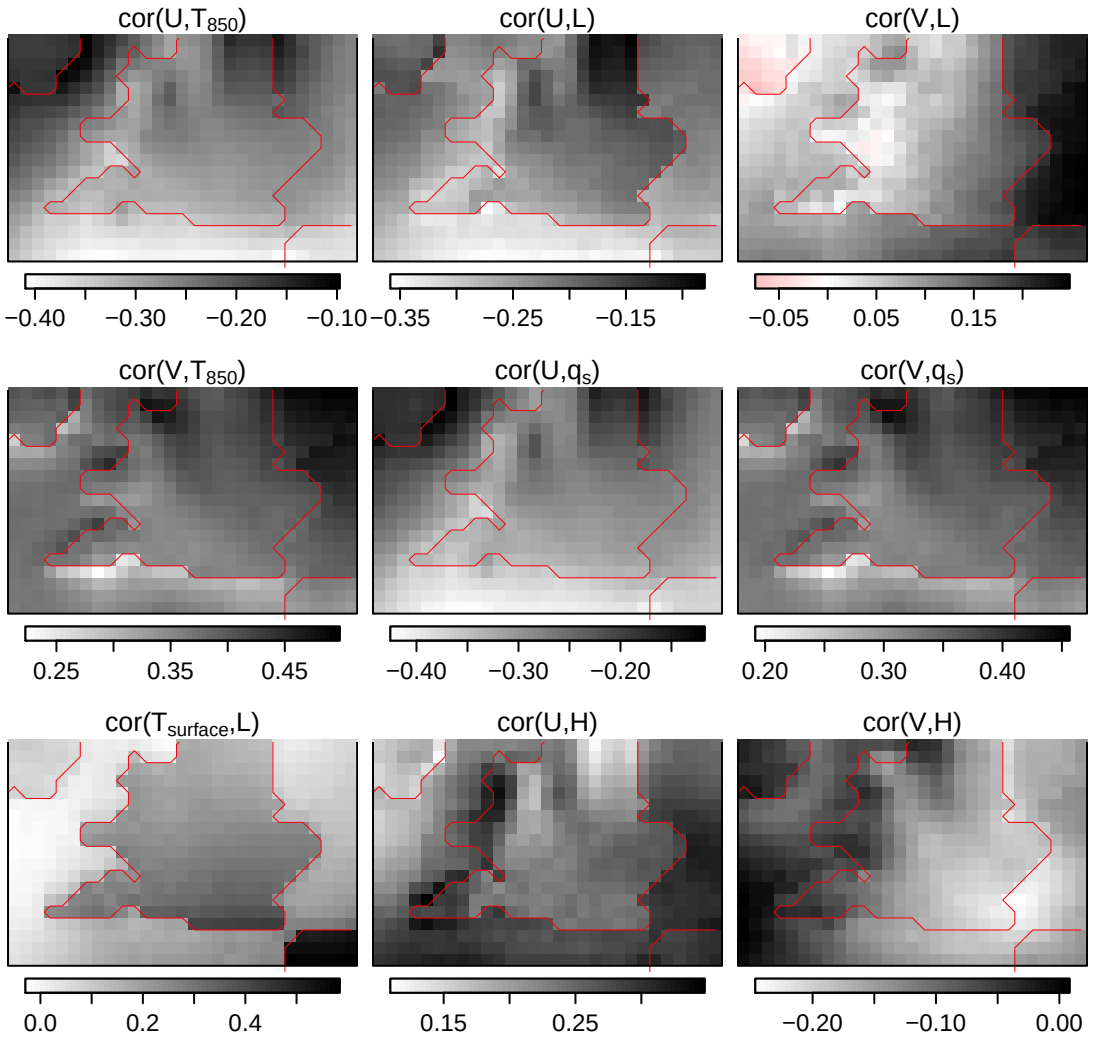

Figure 4.13: Correlations (cor) between the empirical daily-mean lapse rate (L), saturation specific humidity (q_s), relative humidity (H), eastward component of wind (U), northward component of wind (V), temperature (T_{850}) and near-surface air temperature (T_{surface}), all at 850 hPa where applicable, on wet days at the grid-point level across southern Britain. Coastline is in red.

misrepresenting the total variation in the convective precipitation.

Before reporting on the differences between ensemble members, we examine the temporal variation in the convective precipitation within each ensemble member, and the drivers of it. This will develop understanding of the physical process captured by the GLMMs. The bottom-left panel of Figure 4.14 shows that a greater proportion of the total variation in convective precipitation is accounted for by GLMMs over the land (estimated to be up to 37%) than the sea (usually less than 20%). Less of the variation in the convective precipitation is accounted for over the Welsh mountains and along much of the south coast of England than is accounted for over most of the rest of the land. As one might expect, a similar spatial pattern is followed by the variation (bottom-right panel) that is solely due to the temporal variation (FOE($\boldsymbol{X'}$)) in the represented atmospheric conditions.

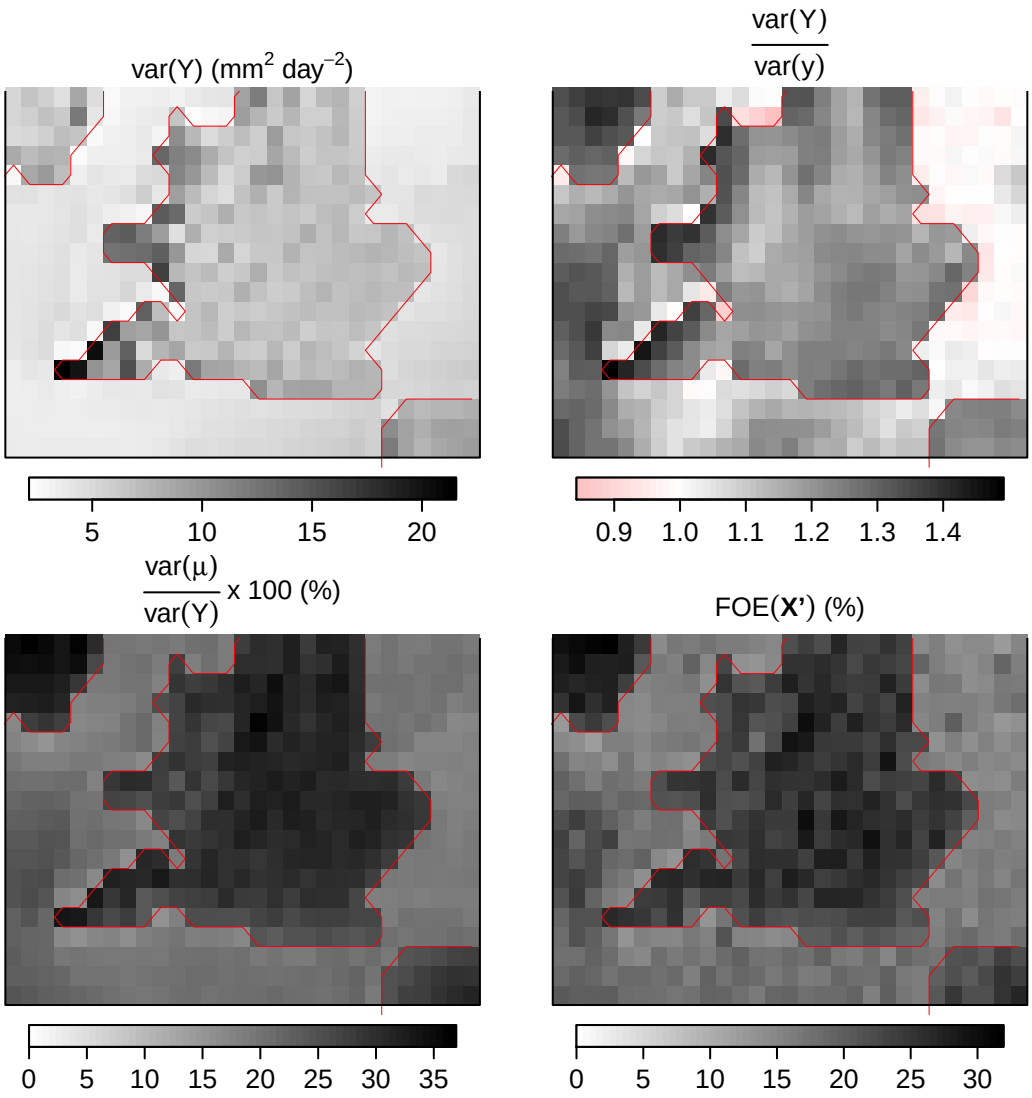

Figure 4.14: Top-left panel: estimated total variation in daily-total summer wet-day convective precipitation (Y) at the grid-point level over southern Britain. Top-right panel: ratio of var(Y) to the empirical variation (var(y)). Bottom-left panel: estimated percentage of the total variation accounted for by the representation of the physical process and additional precipitation bias. Bottom-right panel: First-order effect (FOE) of the temporal variation in the predictors ($\boldsymbol{X'}$), expressed as a percentage of var(Y). Coastline is in red.

To understand the causes of the temporal variation in convective precipitation, Figure 4.15 displays the importance measures of the temporal components in the three predictors individually. As with the Birmingham grid-point, the figure suggests that the temporal variation in lapse rate (top panels) contributes the most to the variation in convective precipitation over the land of southern Britain. On its own, the temporal variation in saturation specific humidity (centre-right panel) contributes little to the variation in convective precipitation, but its interactions (centre-left panel) with the other inputs can be important. This is the case over much of the land, although even these interactions do not seem to influence convective precipitation along the west and south coasts. This spatial pattern over the land can also be seen in the temporal variation in relative humidity (bottom panels) over the land.

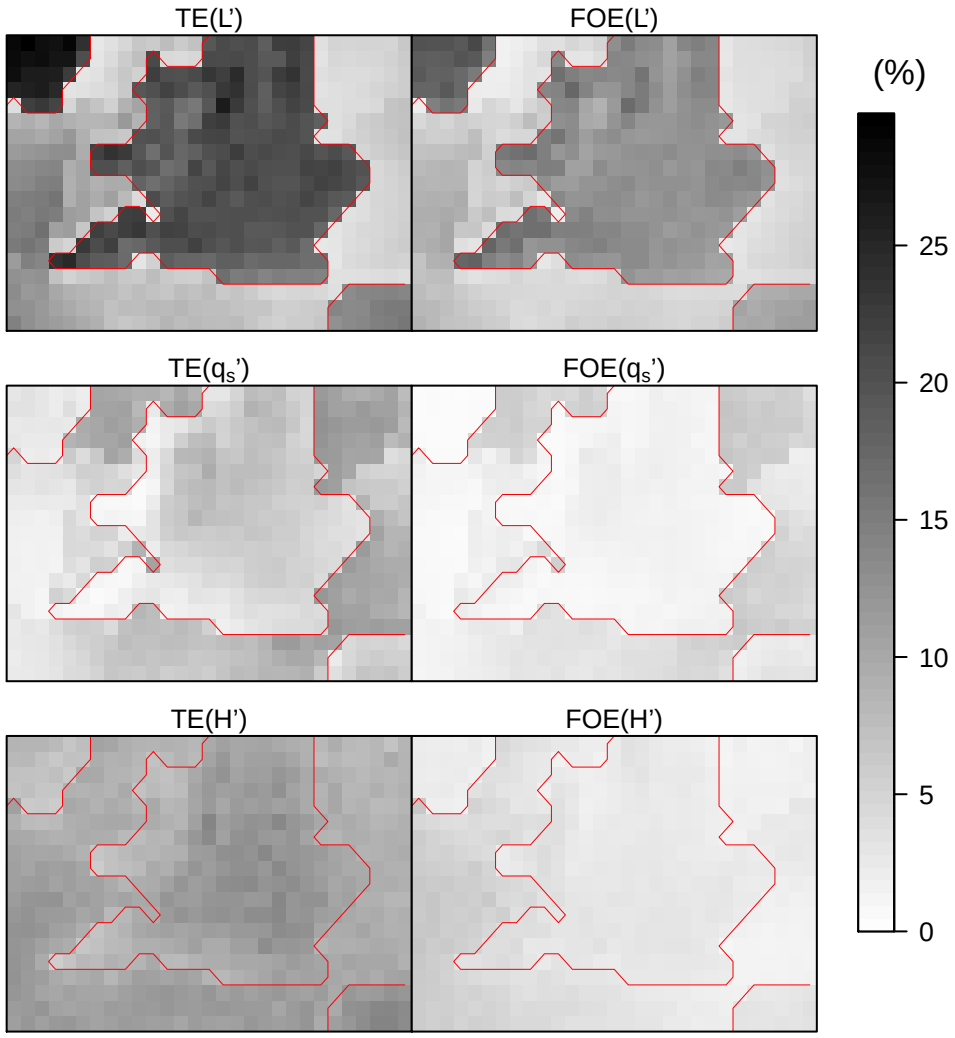

Figure 4.15: Total effects (TE) (on expectation μ) (left panels) and first-order effects (FOE) (right panels) of the temporal variation in lapse rate (L') (top panels), saturation specific humidity (q_s') (centre panels) and in relative humidity (H') (bottom panels) at the grid-point level over southern Britain, expressed as percentages of the total variation (var(Y)) in convective precipitation. Coastline is in red.

The unimportance of the interactions between the predictors along the west and south coasts is likely to be because these coasts are strongly affected by the wind, and the wind has different effects on the predictors (see the six right-most panels of Figure 4.13). Strong westerly winds are associated with low lapse rates, low saturation specific humidities, but high relative humidities along the west coast, and strong southerly winds are associated with the high lapse rate, high saturation specific humidity, and low relative humidity along the south coasts. This emphasizes that the winds affect the drivers of convective precipitation even in the summer.

With respect to the contrast between land and sea, Figure 4.15 shows that the GLMMs account for proportionately less of the variation in convective precipitation

over the sea around Britain than over the land because the variation in lapse rate is less important over the sea. Sea retains heat more efficiently than land, and so warms more slowly each day in the summer. As a result, high lapse rates over the sea are driven less by high surface temperatures (see bottom-left panel of Figure 4.13) and more by southerly winds advecting warmth (recall top-right panel, Figure 4.13) than over land. This could lead to less convection because the convective instability could vertically span less of the atmosphere.

Figure 4.16 focuses on the variation in the convective precipitation that is found to be due to the uncertainty in the parameters of the RCM. The top panels compare such variation that is found when attempting to capture the representation of the fundamental driving physical process of convective precipitation by using our GLMM (top-left panel), to that found when restricting the analysis to the variation across ensemble members in mean precipitation rate by using a random-effects ANOVA (top-right panel). Whereas up to 3% of the total variation in the convective precipitation is found to be due to the differences between ensemble members over southern Britain by the random-effects ANOVA, we find (TE($\boldsymbol{\gamma}, \boldsymbol{\psi}$)) up to 8% by using GLMMs. Moreover as stated previously, this is likely to be underestimated by our GLMMs because much of the variation in the convective precipitation is unaccounted for. This greater variation found using GLMMs is because most is due to variation across ensemble members in the representation of the physical process captured by the predictors and their relationships with convective precipitation, and this variation does not necessarily contribute to variation across ensemble members in the mean precipitation rate. The bottom-left panel highlights the variation confirmed to be due to the representation of this physical process captured by the predictors and their relationships with convective precipitation.

As with the Birmingham grid-point, the bottom-right panel of Figure 4.16 shows that the first-order effect (FOE($\boldsymbol{\psi}$)) of the random effects (approximately) recovers the result from the random-effects ANOVA shown in the top-right panel over the study region. Moreover, each of the three predictors is simulated robustly (not shown) across ensemble members over southern Britain. This implies that there is little variation across ensemble members in the frequency distribution of the represented atmospheric conditions at any grid point in the region. This highlights that the impact of uncertainty in the RCM parameters is largely on the relationship between the predictors and convective precipitation, i.e. given atmospheric conditions can lead to different rates of convective precipitation in different ensemble members. Again, as with the Birmingham grid-point, no random effect (or set of random effects) appears to dominate the variation in convective precipitation due to the differences between the ensemble members, with each of the relationships with the

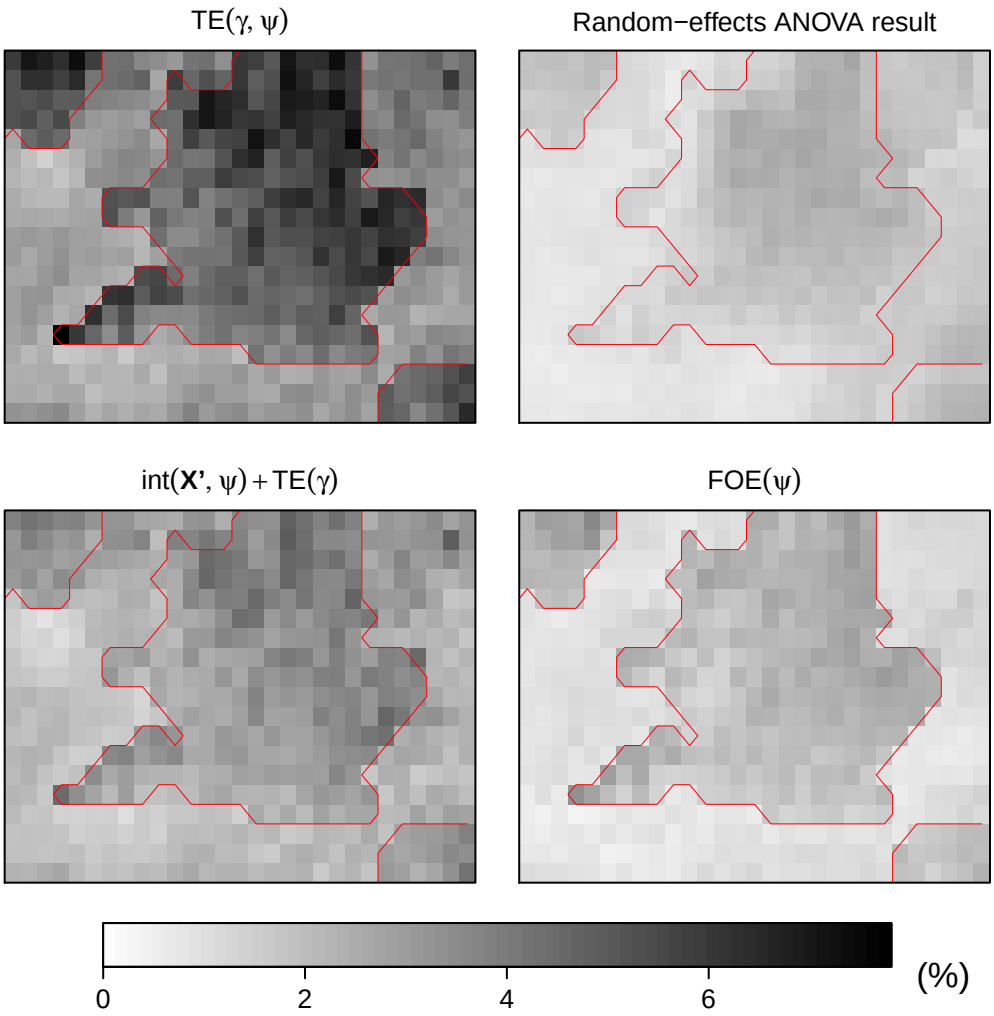

Figure 4.16: Maps concerning the variation in daily-total summer wet-day convective precipitation that is due to the uncertainty in the parameters of the RCM. Coastline is in red. See text for details.

predictors contributing (not shown).

Returning to Figure 4.16 to examine the spatial pattern, it can be seen that convective precipitation is simulated less robustly over land than over sea. This is not surprising because land surface requires parameterizations (recall Section 1.4). The simulation of convective precipitation appears to be the most sensitive to the uncertainty in RCM parameters to the north and east of central England, north-east Wales, and along the north-facing coast of south-west England. We leave a thorough investigation into why the simulation of convective precipitation is least robust in these areas to future work, but provide a possible explanation below. Indeed, determining these reasons would develop understanding of the RCM, which could then lead to the development of climate models such that their simulation of convective precipitation is more robust to uncertainty in the physical parameters of the models. This could also make their projections of convective precipitation in future climates more reliable.

As stated above, the simulation of convective precipitation appears to be most sensitive to uncertainty in RCM parameters to the north and east of central England, north-east Wales, and along the north-facing coast of south-west England. A possible reason for this is that high lapse rate tends to be driven less by high surface temperatures and more by southerly winds advecting warmth in these areas compared to other areas of southern Britain. This can be seen because the correlation between lapse rate and near-surface air temperature (recall bottom-left panel of Figure 4.13) over the land depends on latitude, with greatest correlations in the south, whereas the correlation between lapse rate and the northward component of wind (recall top-right panel, Figure 4.13) is greatest in the eastern half of southern Britain, in north Wales, and along the north-facing coast of south-west England. It may be that defining lapse rate as the temperature at 850 hPa minus the temperature at 500 hPa contributes to the spatial pattern that we find in the sensitivity of the simulation of convective precipitation to the uncertainty in the RCM parameters, but a similar pattern is seen in the random-effects ANOVA result, which does not involve predictors.

4.9 Summary

There is considerable temporal variability on the daily scale in summer convective precipitation over southern Britain. Relative to this temporal variability, perturbing the parameters of HadRM3 results in little variation in the mean rate of convective precipitation (i.e. in precipitation bias). Specifically, up to only 3% of the total variation in the convective precipitation in the ensemble is due to the variation in precipitation bias as a result of the uncertainty in the RCM parameters. This would suggest that perhaps the simulation of convective precipitation is robust to the uncertainty in the parameters of the RCM, but restricting an analysis to precipitation bias cannot quantify the extent to which convective precipitation is simulated robustly in any general sense. This is because for given atmospheric conditions, perturbing the parameters could lead the RCM to simulate different amounts of precipitation, and this will not necessarily be seen in the mean precipitation rates.

The driving physical process of daily-total convective precipitation on a local scale is that moist air becomes convectively unstable, rises enough for the moisture to condense, and this condensed moisture precipitates out. By attempting to capture the representation of this physical process in HadRM3 with a GLMM, we find up to 8% of the total variation in the convective precipitation to be due to the uncertainty in the RCM parameters. Moreover, we consider this to be underestimated because we were unable to capture much of the variation in the precipitation, partly due

to not accounting for the temporal variation in the winds. The 8% of the total variation in convective precipitation that we find can be due to perturbing the RCM parameters is more than twice the 3% found when only analysing precipitation bias, and reveals that most of the variation across ensemble members in the simulation of convective precipitation cannot be detected in its mean on this scale. We find that the simulation of convective precipitation may tend to be more sensitive to uncertainty in the RCM parameters where high lapse rate is driven less by high surface temperature and more by southerly winds advecting warmth, relative to other areas of land, although this requires further investigation.

In addition to attempting to quantify the extent to which summer convective precipitation events are simulated robustly over southern Britain, we find that the temporal variation in lapse rate is a more important driver of convective precipitation over land than is the variation in saturation specific humidity or in relative humidity. We find that the extent to which the variation in these humidity variables drives convective precipitation over land depends on the proximity to the windward coasts. Their variation is relatively unimportant to the precipitation along the west and south coasts of southern Britain. Over the sea, the variation in lapse rate is a less important driver of convective precipitation than it is over land, but the variation in saturation specific humidity is often more important over sea than land. We find that almost no variation is due to the uncertainty in the parameters of the RCM in the simulation of daily-mean lapse rate, saturation specific humidity and relative humidity at the gird-point level over southern Britain.

4.10 Discussion

We analysed daily-total summer wet-day convective precipitation in an RCM perturbed physics ensemble at the grid-point level over southern Britain. The proportion of the total variation (up to 8%) in this precipitation that we found to be due to uncertainty in the parameters of the RCM may appear small, but it is relative to the substantial variation in precipitation on the daily scale. Over the Birmingham grid-point that we used to demonstrate our analysis in detail, for example, we found 6% of the total variation in the convective precipitation to be due to the uncertainty in RCM parameters. We estimated the total variation in convective precipitation to be 6.66 mm^2 day^{-2} (the empirical variance was 5.46 mm^2 day^{-2}). In absolute terms, therefore, we estimate that 0.40 mm^2 day^{-2} of the variation in convective precipitation is due to the uncertainty in the RCM parameters. Expressing this in units that are more relatable to the real world gives the estimated standard deviation due to the uncertainty in the RCM parameters, and this is 0.63 mm day^{-1}. Converting our estimated total variation in the convective precipitation to these units gives

2.58 mm day^{-1}, which is the average deviation from the (ensemble) mean daily-total summer wet-day convective precipitation (i.e. the estimated standard deviation of the convective precipitation). The contribution to this of 0.63 mm day^{-1} due to the uncertainty in the parameters of the RCM is, we suggest, significant.

By relating predictor variables to convective precipitation in the perturbed physics ensemble of HadRM3, we have shown that much of the variation across ensemble members in convective precipitation on a local scale is because of variation across the members in the relationship between the represented atmospheric conditions and the precipitation. Moreover we have shown that this will not necessarily be detectable from the variation across ensemble members in precipitation bias (mean precipitation rate). This emphasizes the need to use GLMMs to capture the representation of driving physical processes of precipitation when attempting to quantify the total variation in the precipitation that is due to the differences between climate models in any general sense. It is the representation of such physical processes that determines the extent to which precipitation events are simulated robustly across climate models.

As discussed in Subsection 1.4.1, it is well established that convective precipitation is difficult to represent in climate models largely as a result of the fine spatial scales on which it occurs, because this results in the driving physical processes of convective precipitation being represented by subgrid-scale parameterizations. Over the land of southern Britain we found that there is usually enough moisture available in the atmosphere to mean that the temporal variation in lapse rate is the main driver of convective precipitation. Our results suggest that the simulation of convective precipitation may be most sensitive to the uncertainty in the RCM parameters over land where it is relatively common for high lapse rate to be caused by southerly wind advecting warm air to the area, rather than by high surface temperatures. This suggests that the representation of convection may be more sensitive to the uncertainty in the RCM parameters above the boundary layer compared to when the convection is triggered by high surface temperatures. This is the case in the east of southern Britain, for example, where we found that the proportion of variation in convective precipitation that is due uncertainty in RCM parameters is typically 5 to 7%. This is typically 3 to 4% near the south coast, where high surface temperatures are prominent. This requires further investigation, however, because determining the causes of disagreement between ensemble members is key to the understanding and development of climate models.

The aim of relating lapse rate, saturation specific humidity and relative humidity to convective precipitation was to capture the representation of the driving physical

process of the convective precipitation at the grid-point level on the daily scale. We therefore aimed to account for most of the variation in the precipitation in the ensemble. However we only accounted for up to 37% at any grid point over the study region, finding up to 8% to be due to uncertainty in the parameters of the RCM. The result of only accounting for up to 37% of the total variation in convective precipitation in the ensemble is that the proportion that we find to be due to the uncertainty in RCM parameters is likely to be an underestimate of the total influence of perturbing the RCM parameters on convective precipitation. Therefore our analysis only begins to quantify the extent to which the simulation of convective precipitation is robust to the uncertainty in the parameters of climate models in any general sense. Further work is to capture more of the variation in the representation of the driving physical process of convective precipitation events in the ensemble, perhaps by correcting for the winds, as well as to use larger ensembles which include different RCMs. Moreover the influence of the differences between climate models on the occurrence of wet days of convective precipitation could be investigated.

In addition to the future work outlined above, further insight into why precipitation varies across perturbed physics ensembles can be gained by investigating the influence of perturbing each individual parameter of the climate model. Therefore in the next chapter we focus on how the values of individual climate model parameters can be incorporated into a GLMM, and how the extent to which perturbing each climate model parameter influences the simulation of precipitation can be quantified.

5 RCM parameters and summer precipitation

Aim: To demonstrate how to quantify the extent to which uncertainty in each parameter of an RCM contributes to the variation in its simulation of summer precipitation.

5.1 Introduction

In Chapter 4 we attempted to quantify the extent to which, and understand why, convective precipitation varies across and within a perturbed physics ensemble of the regional climate model (RCM) HadRM3 (see Section 4.2 and references therein). However we fell short of analysing the influence of each individual parameter of HadRM3 on the convective precipitation. In this chapter we demonstrate how to quantify the extent to which the uncertainty in each individual parameter of the RCM contributes to the variation in its simulation of precipitation. To do this we use the same ensemble as in Chapter 4, and the same summer months (June, July and August) of the same period of time (years 1960-1989). Recall from Section 4.2 that the ensemble provides gridded data with a spatial resolution of 25 km, and that the ensemble is from the HadRM3-PPE-UK experiment (see Met Office Hadley Centre, 2008) at the UK Met Office Hadley Centre, which covered the wider European area after dynamically downscaling a general circulation model (GCM) before extracting the British Isles region. Also recall that the ensemble has eleven members, and that a total of 33 parameters are perturbed. Note that the extent to which perturbing each individual parameter of the RCM contributes to the variation in its simulation of precipitation depends on the extent to which each parameter is perturbed in the ensemble used, and so we assume that the perturbations are representative of the uncertainty in each RCM parameter.

Our focus in Chapter 4 was on the cumulative effect of all of the uncertainty sampled by perturbing the RCM parameters, and on the driving physical process of (summer) convective precipitation on the local scale, using daily total/mean data at the grid-point level over southern Britain. Aggregating the data up to a coarser temporal resolution would, therefore, have been counter-productive because detail would have been lost in the physical process of interest. Additionally, restricting the precipitation to only convective precipitation allowed us to focus on this physical process and more easily diagnose the causes of the variation within the ensemble members. In this chapter, however, we are interested in the individual effect of the uncertainty in each RCM parameter on the simulation of summer precipitation more generally than focusing on a specific physical process because knowing the role

within the RCM of each parameter automatically provides insight into the causes of their effects on the precipitation (e.g. some RCM parameter could be known to control the representation of convection in the RCM). The individual effect of the uncertainty in each RCM parameter on the representation of specific physical processes may also be of interest, but we leave this to future analyses.

As a result of analysing the individual effect of the uncertainty in each RCM parameter on the simulation of summer precipitation generally, in this chapter we analyse all summer precipitation (not just convective precipitation), broaden the study area out to all of the British Isles (not just southern Britain), and aggregate the precipitation totals up from the daily scale to the seasonal scale. As such, we analyse summer-total precipitation at the grid-point level over the British Isles in this chapter. Aggregating up to the seasonal scale reduces the temporal variation and increases the signal from perturbing each parameter of HadRM3. As stated above, the cost of this aggregation is that we lose detail on the daily representation of physical processes, but increasing the signal of the effect of perturbing individual parameters can be seen as an advantage for this application. Additionally, an advantage of analysing all summer precipitation in this chapter is that totals of all precipitation (large-scale and convective) in climate models are more relatable to precipitation totals in the real world.

On the seasonal scale, perhaps the most important driver of precipitation over the British Isles is air pressure (see Briffa et al., 1986). High pressure in European (including British) summers is associated with dry sunny conditions, and sometimes heat waves (e.g. see Sutton and Hodson, 2005, and Della-Marta et al., 2007). In general, masses of air advect from areas of high pressure to areas of low pressure within the boundary layer (see page 294 of Wallace and Hobbs, 2006, and the rest of their Chapter 7 for further detail; also see Garratt and Physick, 1983, and Johnson and Hamilton, 1988). The result is that air is forced to ascend at areas of low pressure due to the convergence of low-level atmospheric flow, potentially leading to the influx of moisture and to precipitation, whereas there is relatively little ascent at areas of high pressure due to the divergence of air from these areas.

In this chapter we relate summer-mean surface air pressure to summer-total precipitation over the British Isles using a statistical model at each grid point, and incorporate the variation in each individual parameter of HadRM3 that is perturbed in the ensemble into the statistical modelling. Our focus is on this variation across ensemble members in the parameters of the RCM, rather than the variation in surface pressure. Therefore including a predictor is not necessary to assess the influence of perturbing each RCM parameter on the mean precipitation rate in each ensemble

member, but we use surface pressure as a predictor here because its interaction with each RCM parameter may contribute to the variation in precipitation. In other words, we use pressure to summarize the temporal variation in the atmospheric conditions of each summer within each ensemble member, and the perturbations in the RCM parameters could result in given atmospheric conditions leading to different amounts of precipitation.

Surface pressure is used as a single predictor in this application. An alternative to this would be to use different directional components (e.g. the northward component and the eastward component) of pressure gradients, or perhaps include wind components, to capture the effect on precipitation of air masses being advected from different directions (which could affect the temperature of the air, for example). We do not feel the need to do this in this chapter, however, because as we show in the next section, using surface pressure as a single predictor summarizes the atmospheric conditions well enough to correlate strongly and negatively (as expected) with the precipitation.

The aim of this application is to demonstrate how to quantify the extent to which uncertainty in each parameter of an RCM contributes to the variation in its simulation of summer precipitation. If the importance of each RCM parameter to precipitation can be quantified, then (for example) the most important RCM parameters can be determined, and reducing the uncertainty in their optimal values can perhaps be made the focus when developing climate models. Recall from Subsection 1.4.2 that some climate model parameters (e.g. the entrainment coefficient and the ice-fall speed parameter) have already been identified as important to precipitation, but importance measures are able to quantify (and therefore rank) the importance of each RCM parameter formally, which cannot be achieved by methods such as examining exploratory plots. Moreover, by using importance measures to quantify the importance of each RCM parameter at each grid point over a region of interest, maps of the importance measures can be produced to examine the spatial pattern. Determining the causes of such spatial patterns could further reveal how the simulation of precipitation is affected by the uncertainty in each RCM parameter.

We stated above that 33 parameters of HadRM3 are perturbed in the ensemble. As we detail in the next section, however, not all of these RCM parameters are perturbed independently. This is because the perturbations in some parameters have perfect correlation in the ensemble, and means that only 27 of the parameters are perturbed independently. Even given this, there are more parameters (27) of HadRM3 that are perturbed in the ensemble than there are ensemble members (eleven), and so there would not be enough degrees of freedom to include all of the

RCM parameters together in the same statistical model. Therefore to incorporate the variation in each individual perturbed parameter into the statistical modelling, we include each RCM parameter in the statistical modelling once at a time at each grid point, i.e. after analysing one RCM parameter we remove it from the statistical model, re-fit the model with a different RCM parameter included, and repeat this for all 27 independently perturbed parameters. We detail this procedure later, but note that it is only necessary because of the relatively few ensemble members available; if there were more ensemble members than perturbed parameters then all of the RCM parameters could be included in a single statistical model.

To estimate the variation in summer precipitation that can be accounted for by the perturbations in each of the RCM parameters perturbed in the ensemble, we use the methodology of Chapter 2 to estimate the importance measures of each RCM parameter. We do this at each grid point over the British Isles and examine the spatial patterns. The aim of quantifying the importance of uncertainty in individual RCM parameters in this way is to aid the development of climate models: if the important sources of uncertainty are to be reduced in climate models, the important sources must be identified first. As we show in the next section, there are not enough members in the ensemble described above to disentangle the effects of perturbing individual RCM parameters, but in this chapter we provide a method to do this with larger ensembles by demonstrating it with the perturbed physics ensemble available to us.

5.2 Exploratory data analysis

Plotting the data to inform the statistical modelling, Figure 5.1 displays the relationship between surface pressure (P) and summer-total precipitation (Y) in each ensemble member at a single grid point over Birmingham. The relationship is negative in each ensemble member, as expected. There is some variation across the members in the strength of the relationship, such that (for example) a low-pressure summer of approximately 995 hPa would lead to an expected precipitation total of 250 mm in one ensemble member (panel of row 2 column 1) and 300 mm in another (row 2 column 2). The variation in surface pressure seems more important to the precipitation, however, with summers of high pressure (1005 hPa) leading to less than 50 mm of precipitation in some ensemble members (e.g. row 1 column 3, and row 3 column 2). There appears to be some non-linearity in the relationship in some ensemble members, but assuming linearity would approximate the relationship well enough for our analysis. Similarly, there appears to be a decrease with pressure in the variation in precipitation in some ensemble members, and some variation across the members in the tightness of the points to the curves, but assuming a constant

residual term could provide an adequate approximation for our application. Note that the relationships between surface pressure and summer-total precipitation appear similar (approximately linear, etc) at other grid points (not shown), suggesting that the plots of Figure 5.1 are representative of the study area generally.

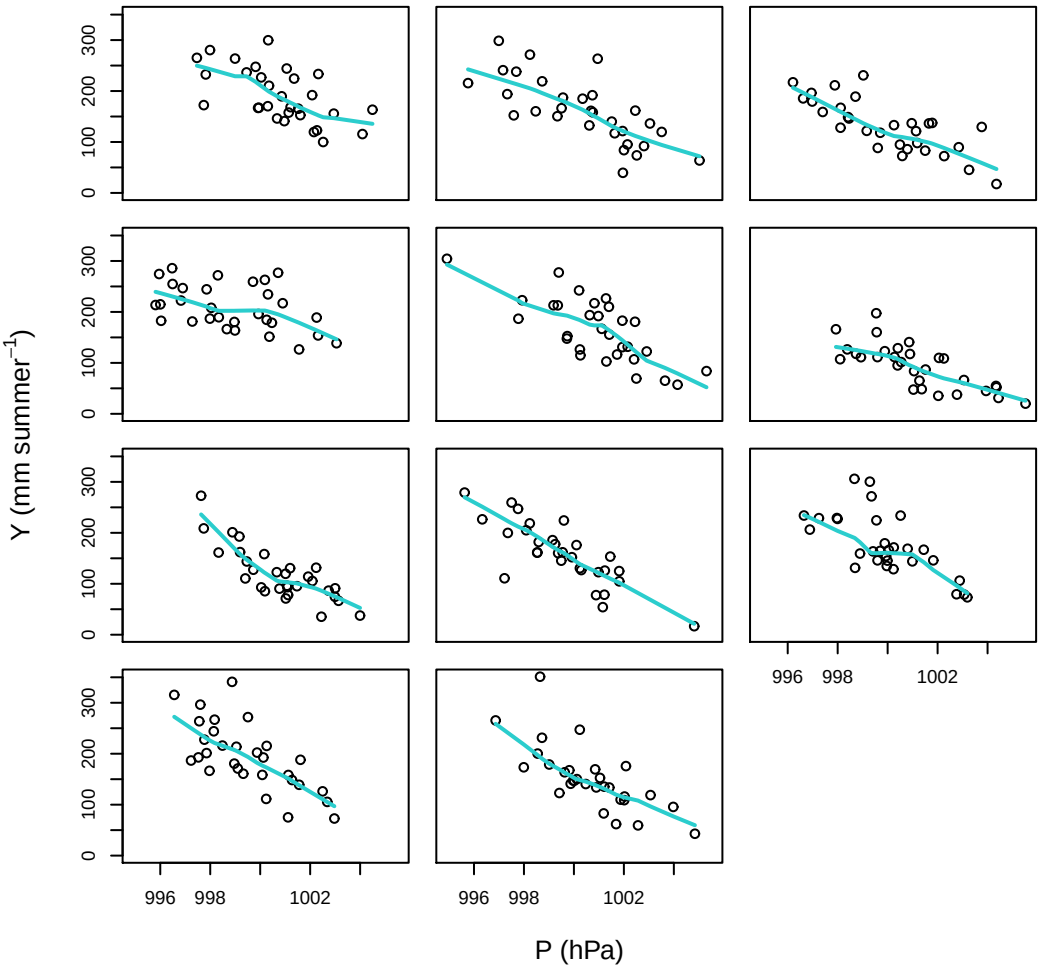

Figure 5.1: Summer-total precipitation (Y) against summer-mean surface pressure (P) in each ensemble member at a single grid point over Birmingham. Curves are local polynomial smoothers. The control run is top left.

The box plots of Figure 5.2 show that there is considerable variation across ensemble members in the distributions of summer-total precipitation and of summer-mean surface pressure, relative to their temporal variation within each member. These distributions (negatively) correlate, although much of the variation across the members in the distribution of precipitation cannot be explained by that across the members in the distribution of the surface pressure.

The individual parameters of HadRM3 that are perturbed in the ensemble are listed, labelled and described in Table 5.1, and Figure 5.3 shows how the perturbations of each of these RCM parameters relates to the summer precipitation in the ensemble

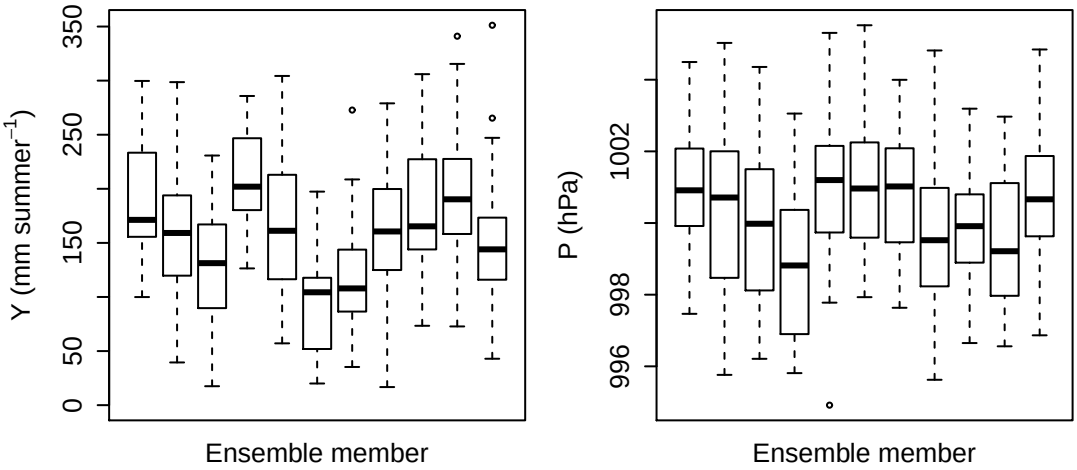

Figure 5.2: Box plots of summer-total precipitation (Y) (left panel) and summer-mean surface pressure (P) (right panel) by ensemble member at a single grid point over Birmingham. The control run is on the left of each panel.

at the Birmingham grid-point. Four parameters (rhparam, sw_absn, canopy and sto_res) are either switched on or off (panels of row 2 column 2, row 4 column 2, row 5 column 6, and row 6 column 1 of Figure 5.3). This contrasts to the timescale for destruction of the convective available potential energy (CAPE) parameter (Cape), which is switched off in six of the eleven ensemble members but takes different values when switched on in the others. Therefore the relationship between Cape and the precipitation is shown by two panels (row 2 columns 4 and 5) of Figure 5.3. Similarly, the convective updraught factor (cnv_upd) and convective anvil shape factor (anvil) are also switched off in (the same) six ensemble members (for which Cape is also switched off). Therefore three panels (row 2 column 6, and row 3 columns 1 and 2) of Figure 5.3 are associated with these two parameters.

It appears from Figure 5.3 that increasing some parameters has a positive effect (e.g. the ice-fall speed, VF1, panel of row 1 column 2) on the amount of summer precipitation over Birmingham, whereas increasing some has a negative effect (e.g. the ice particle size, Icesize, row 3 column 3), and increasing others has little effect (e.g. the gravity wave drag start level, gw_lev, row 6 column 5). As one might expect, the relationships are often inverted when the RCM parameters are related to surface pressure (see Figure 5.4). It can also be seen from Figures 5.3 and 5.4 that the relationships between the perturbations and the pressure are often weaker than those between the perturbations and the precipitation. As stated in Section 4.2, the ensemble samples the uncertainty in the physical parameters of the RCM (HadRM3), with the boundary conditions driven by a GCM. Air pressure is a large-scale variable, often influenced by the movement of large masses of air, and so the weaker relationships between perturbing the RCM parameters and the pressure is perhaps because its simulation is largely driven by the GCM, which is common across the

	Parameter	Label	Description
Large-scale cloud			
1	RH_{crit}	RHCrit	Threshold of relative humidity for cloud formation
2	Ice-fall speed, V_{f1}	VF1	Ice particle fall speed
3	C_w	CW_land	Cloud droplet to rain conversion threshold over land
4	C_w	CW_sea	Cloud droplet to rain conversion threshold over sea
5	C_t	CT	Cloud droplet to rain conversion rate
6	Cloud fraction at grid-box saturation boundary layer	eacfbl	Cloud cover calculation
7	Cloud fraction at grid-box saturation free troposphere value	eacftrp	
8	Flow-dependent RH_{crit} scheme on/off	rhparam	Parameterization of RH_{crit} in terms of local variance of cloud water
Convection			
9	Entrainment coefficient	Ent	Rate of mixing between environmental air and convective plume
10	Timescale for destruction of CAPE	Cape	Intensity of convective mass flux
11	Convective updraught factor	cnv_upd	Fraction of convective cloud in which updraught occurs
12	Convective anvil shape factor	anvil	Shape of convective cloud
Radiation			
13	Ice particle size (μm)	Icesize	Effective radius of cloud ice spheres
14	Non-spherical ice particles stratospheric shortwave	s_sph_sw	Accounts for effect of non-spherical ice particles
15	Non-spherical ice particles stratospheric longwave	s_sph_lw	
16	Non-spherical ice particles convective shortwave	c_sph_sw	
17	Non-spherical ice particles convective longwave	c_sph_lw	
18	Shortwave water vapour continuum absorption on/off	sw_absn	Accounts for shortwave absorption due to the self-broadened continuum of water vapour
Sea ice			
19	Sea ice albedo	MinSIA	Dependence of sea ice albedo on temperature
20	Ocean-ice diffusion coefficient	oi_diff	Ocean to heat transfer
21	Sea ice minimum temperature	ice_tr	
Boundary layer and surface processes			
22	Boundary layer flux profile parameter, G_0	flux_g0	Functions used to determine stability dependence of turbulent mixing coefficients
23	Asymptotic neutral mixing length parameter, λ	lambda	Calculation of turbulent mixing coefficients
24	Charnock constant	Charnoc	Roughness lengths and surface fluxes over sea
25	Free convective roughness length over sea	cnv_rl	Surface fluxes over tropical oceans
26	Roughness length, z_0	f_rough	Surfaces fluxes over areas containing forest
27	Number of soil levels accessible for transpiration	soillev	Root depths
28	Surface-canopy decoupling scheme on/off	canopy	Accounts for effect of vegetation canopy on surface energy budget
29	Stomatal conductance response to ΔCO_2 on/off	sto_res	Variance of stomatal conductance with carbon dioxide concentration
Dynamics			
30	Dynamical diffusion e-folding time	dyndiff	Diffusion coefficients for heat, momentum and moisture
31	Surface gravity wave constant	k_gwd	Magnitude of hydrostatic gravity wave stress
32	Trapped lee wave constant	k_lee	Magnitude of non-hydrostatic gravity wave stress
33	Gravity wave drag start level	gw_lev	Lowest model level at which drag is applied

Table 5.1: List of perturbed parameters of HadRM3, their labels and descriptions. Labels are taken from Pope et al. (2011) and descriptions from Barnett et al. (2006).

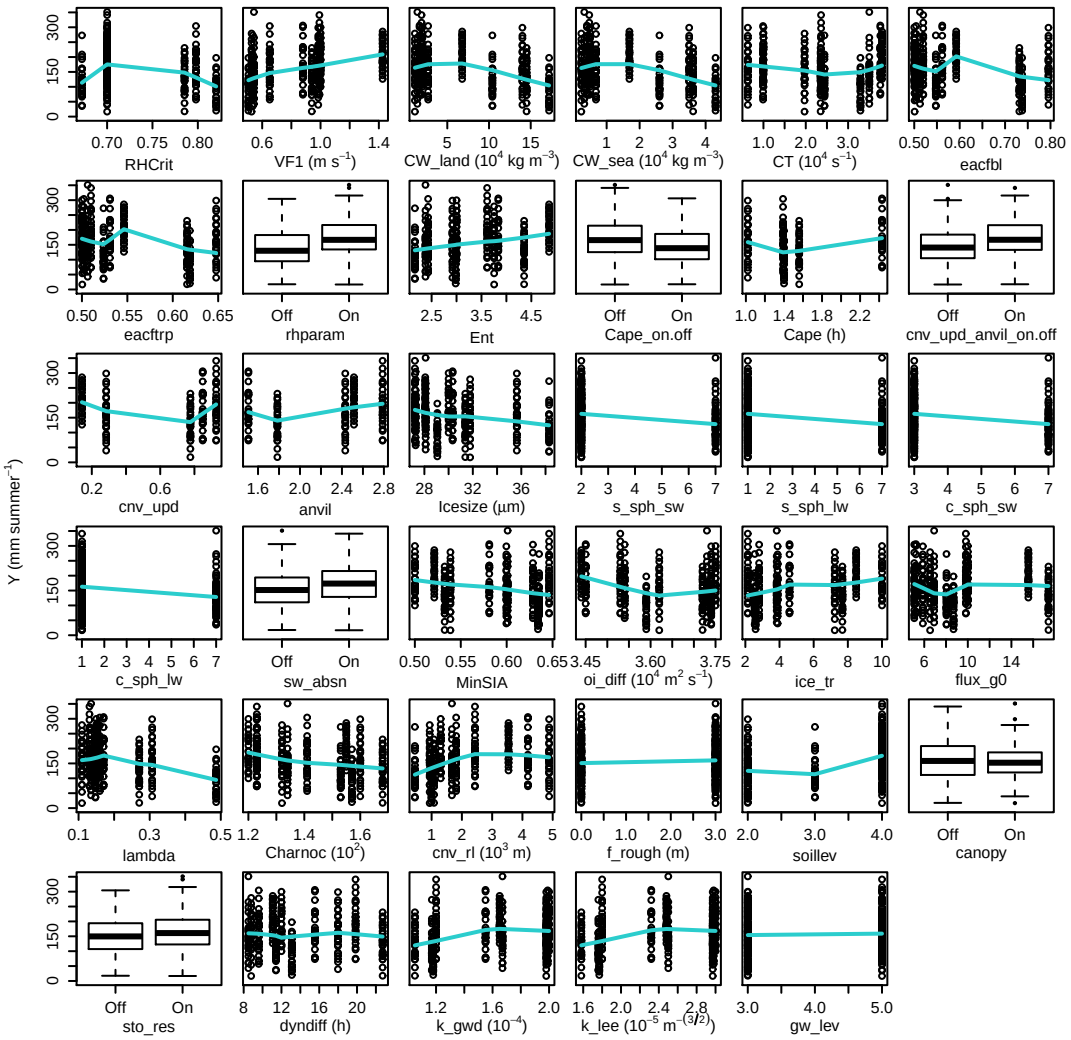

Figure 5.3: Summer-total precipitation (Y) against perturbed parameters (see Table 5.1 for a description of the labels) of HadRM3 at the Birmingham grid-point. Curves are local polynomial smoothers.

ensemble members.

As noted in Section 5.1, a total of 33 parameters are perturbed in HadRM3, although not all of the 33 parameters are done so independently. The cloud droplet to rain conversion threshold over land (CW_land) and that over sea (CW_sea) are perturbed together (i.e. have perfect correlation across the ensemble members), and so we group them together by only using one in the statistical modelling and labelling it CW. Similarly, the cloud fraction at grid-box saturation boundary layer (eacfbl) and the cloud fraction at grid-box saturation free troposphere value (eacftrp) are perturbed together; we label this group eacf. The non-spherical ice particles stratospheric shortwave (s_sph_sw), the non-spherical ice particles stratospheric longwave (s_sph_lw), the non-spherical ice particles convective shortwave (c_sph_sw), and the non-spherical ice particles convective longwave (c_sph_lw) parameters are perturbed together; we group these and label them ice_type. The only other group of param-

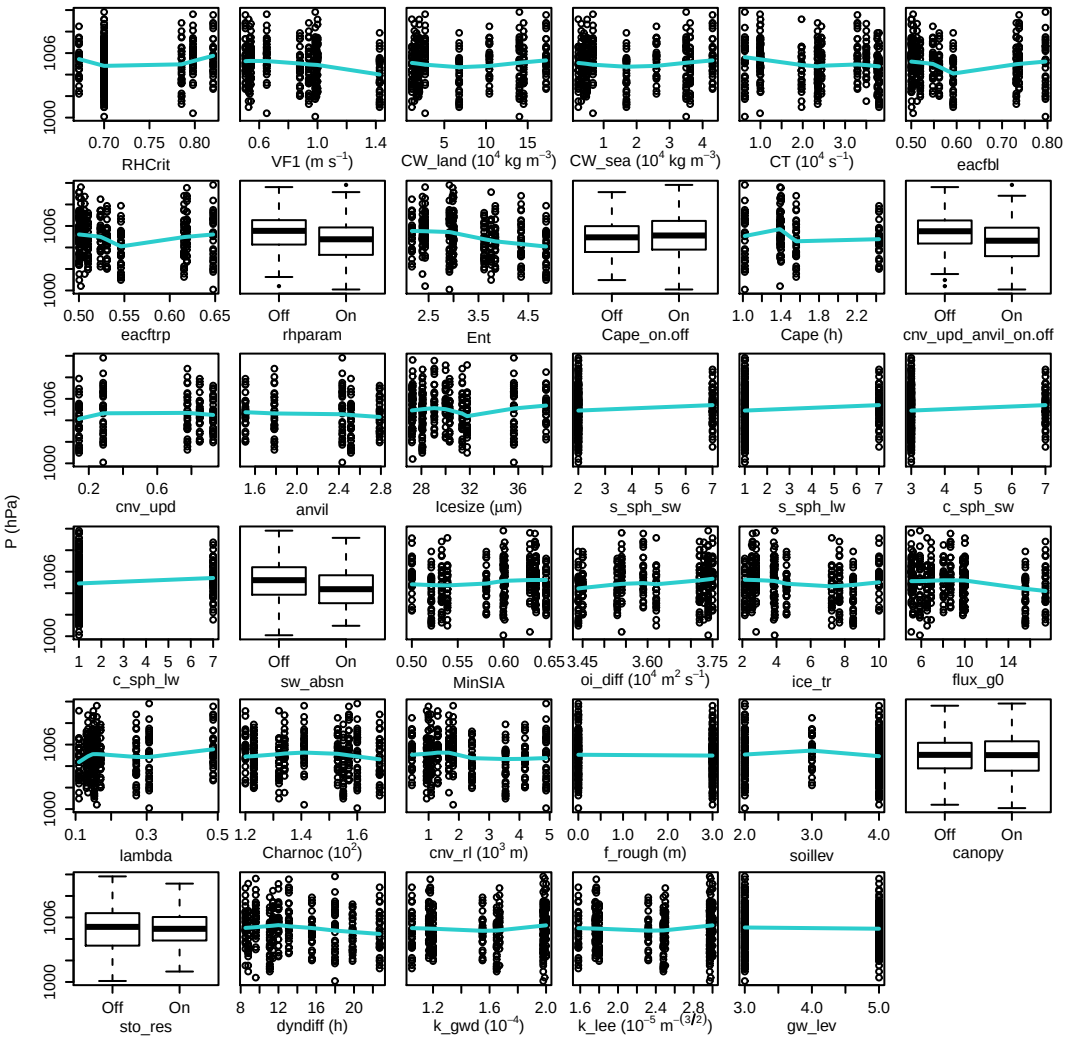

Figure 5.4: Summer-mean surface pressure (P) against perturbed parameters (see Table 5.1 for a description of the labels) of HadRM3 at the Birmingham grid-point. Curves are local polynomial smoothers.

eters that are perturbed together is the surface gravity wave constant (k_gwd) and the trapped lee wave constant (k_lee), which we label together as k. As a result of these groupings, a total of 27 parameters of HadRM3 are perturbed independently across the ensemble members.

Even between the 27 independently perturbed parameters of the RCM, however, there are strong correlations (see Table 5.2). These correlations occur because there are too few ensemble members (eleven) to perturb each RCM parameter whilst holding all of the others at nominated values (and this would not test any interactions between parameters in the ensemble), rather than as a result of any dependence between the uncertainty in the parameters (see Barnett et al., 2006, for discussion on the design of perturbed physics ensembles). The implication of the strong correlation between the perturbations is that the variation in the precipitation that can be accounted for by the perturbations in a parameter is non-orthogonal with that

accounted for by the perturbations in other parameters. For example the first row of Table 5.2 highlights that the perturbations in the threshold of relative humidity for cloud formation (RHCrit) are correlated in the ensemble with whether the flow-dependent RH_{crit} scheme is switched on or off (rhparam), the perturbations in the timescale for destruction of CAPE (Cape), those in the asymptotic neutral mixing length parameter (lambda), and with the perturbations in whether the stomatal conductance response to ΔCO_2 is switched on or off (sto_res). Therefore the relatively few ensemble members means that it is difficult to disentangle the effect of the perturbations in the threshold of relative humidity for cloud formation (RHCrit) with the effects of the perturbations in these other parameters. As a result, we can quantify the variation in precipitation that can be accounted for by the perturbations in each RCM parameter in the ensemble, but cannot assume that these quantifications generalize to larger or different ensembles. Given the relatively few ensemble members, it is perhaps advantageous to guard against over-fitting by assuming linear relationships in the statistical modelling, despite some non-linear relationships seen in Figure 5.3 (e.g. that between RHCrit and the precipitation, top-left panel).

5.3 Statistical model specification

As stated in the introduction (Section 5.1) to this chapter, we quantify the variation in summer-total precipitation that can be accounted for by the perturbations in each parameter of HadRM3 using a statistical model specification fitted 27 times at each grid point, with a different RCM parameter included each time. We use these statistical models to estimate the importance measures from Chapter 2 of each RCM parameter. If the ensemble consisted of enough members then we could include all of the perturbed parameters in the same statistical model (and consider interactions between them), but we are unable to do this here. Therefore we relate summer-mean surface pressure (P) and a single RCM parameter (R) to summer-total precipitation (Y) in each statistical model at each grid point separately over the British Isles. Based on the exploratory data analysis, we assume linear relationships in the statistical models, and allow the relationship between pressure and precipitation to vary with each RCM parameter by including an interaction term between them.

In Section 5.1 (and in the paragraph above) we stated that we use surface pressure as a predictor, and we examined surface pressure in Section 5.2. More specifically, however, we use the temporal variation (P') in the surface pressure as a predictor in the statistical modelling, such that P' within each ensemble member is centred on the ensemble-mean pressure at each grid point. This means that P' differs from inputs that capture temporal variation in predictors in other chapters because such inputs are centred on zero within each ensemble member in the other chapters, whereas P'

	Parameter	Correlations
Large-scale cloud		
1	RHCrit	sto_res (−0.68), lambda (0.67), rhparam (−0.57), Cape (0.55)
2	VF1	anvil (0.79), Ent (0.69), cnv_rl (0.60), flux_g0 (0.58), ice_tr (0.51), MinSIA (−0.50)
3	CW	soillev (−0.96), eacf (0.90), Charnoc (0.88), lambda (0.84), sto_res (−0.67), f_rough (−0.50)
4	CT	k (−0.51)
5	eacf	soillev (−0.92), CW (0.90), lambda (0.84), Charnoc (0.78), sto_res (−0.70), dyndiff (0.58), f_rough (−0.57), gw_lev (−0.56)
6	rhparam	cnv_rl (0.65), RHCrit (−0.57), canopy (0.56), anvil (0.51)
Convection		
7	Ent	VF1 (0.69), flux_g0 (0.68), anvil (0.64), ice_type (−0.59), MinSIA (−0.59), ice_tr (0.57), oi_diff (−0.52), cnv_upd (0.51)
8	Cape	RHCrit (0.55)
9	cnv_upd	dyndiff (0.83), anvil (0.72), Ent (0.51)
10	anvil	VF1 (0.79), cnv_upd (0.72), cnv_rl (0.71), dyndiff (0.70), Ent (0.64), rhparam (0.51)
Radiation		
11	Icesize	oi_diff (0.56)
12	ice_type	Ent (−0.59)
13	sw_absn	gw_lev (0.52)
Sea ice		
14	MinSIA	ice_tr (−0.99), Ent (−0.59), flux_g0 (−0.51), VF1 (−0.50)
15	oi_diff	Icesize (0.56), Ent (−0.52)
16	ice_tr	MinSIA (−0.99), Ent (0.57), flux_g0 (0.55), VF1 (0.51)
Boundary layer and surface processes		
17	flux_g0	Ent (0.68), VF1 (0.58), ice_tr (0.55), MinSIA (−0.51)
18	lambda	CW (0.84), eacf (0.84), soillev (−0.83), RHCrit (0.67), sto_res (−0.63), Charnoc (0.59), f_rough (−0.57), gw_lev (−0.50)
19	Charnoc	CW (0.88), soillev (−0.81), eacf (0.78), sto_res (−0.78), lambda (0.59)
20	cnv_rl	anvil (0.71), rhparam (0.65), VF1 (0.60)
21	f_rough	dyndiff (−0.70), soillev (0.58), eacf (−0.57), lambda (−0.57), CW (−0.50)
22	soillev	CW (−0.96), eacf (−0.92), lambda (−0.83), Charnoc (−0.81), f_rough (0.58), sto_res (0.58), dyndiff (−0.53)
23	canopy	rhparam (0.56), k (0.50)
24	sto_res	Charnoc (−0.78), eacf (−0.70), RHCrit (−0.68), CW (−0.67), lambda (−0.63), soillev (0.58)
Dynamics		
25	dyndiff	cnv_upd (0.83), anvil (0.70), f_rough (−0.70), eacf (0.58), soillev (−0.53), k (−0.52)
26	k	dyndiff (−0.52), CT (−0.51), canopy (0.50)
27	gw_lev	eacf (−0.56), sw_absn (0.52), lambda (−0.50)

Table 5.2: Correlations of magnitude 0.5 or greater between the perturbed parameters of HadRM3 in the ensemble, e.g. the correlation between the perturbations of RHCrit (first row) and those of sto_res is −0.68 in the ensemble. When parameters are switched off they are set to zero when computing the correlations.

is centred on the ensemble-mean pressure within each member in this application. We detail why we only use the temporal variation in the pressure after we define the specification of the statistical model below. Each statistical model corrects for the effect of whichever RCM parameter R is included, and uses a random effect ($\psi^{(0)}$) to capture the remaining variation across ensemble members in the intercept, and a random effect ($\psi^{(1)}$) to capture the remaining variation across the members in the relationship between the surface pressure and precipitation.

Indexing the year of the summer by i and the ensemble member by j, we use a general linear mixed model (LMM) (see Breslow and Clayton, 1993) with specification

$$\begin{aligned} Y_{ij} \mid \left(\psi_j^{(0)}, \psi_j^{(1)}, P'_{ij}, R_j\right) &\sim \text{Normal}(\mu_{ij}, \sigma^2) \\ \mu_{ij} &= \beta_0 + \beta_1 P'_{ij} + \beta_2 R_j + \beta_3 P'_{ij} R_j + \psi_j^{(0)} + \psi_j^{(1)} P'_{ij} \\ \left(\psi_j^{(0)}, \psi_j^{(1)}\right) &\sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}) \ , \end{aligned}$$

where μ denotes the fitted expectation of Y, σ^2 the (constant) residual variation, and $\mathbf{\Sigma}$ the variance-covariance matrix of the random effects. Another way of viewing this statistical model specification is that the random effects are allowed to vary with the perturbations in the RCM parameter. To emphasize this, an equivalent way of writing the LMM is

$$\begin{aligned} \mu_{ij} &= \beta_0 + \beta_1 P'_{ij} + \psi_j^{(0)} + \psi_j^{(1)} P'_{ij} \\ \begin{pmatrix} \psi_j^{(0)} \\ \psi_j^{(1)} \end{pmatrix} &\sim \text{MVN}\begin{pmatrix} \beta_2 R_j \\ \beta_3 R_j \end{pmatrix}, \mathbf{\Sigma} \Bigg) \ . \end{aligned}$$

In this way of writing the LMM it is perhaps clearer that the random effects allow for variation across ensemble members that cannot be captured by the linear effect of perturbing the RCM parameter. Indeed, the random effects would not be necessary if more ensemble members were present and all of the RCM parameters could be included in a single LMM. The random effects are necessary here because we only include one RCM parameter at a time in the statistical modelling, and we assume that the perturbations in any single RCM parameter cannot individually account for all of the variation across ensemble members in the simulation of precipitation. In a given LMM, therefore, the random effects capture the variation in precipitation due to the perturbations in the RCM parameters that are not included (i.e. in all of the processes controlled by these excluded RCM parameters). Representing all of the RCM parameters in each LMM in this way (one as a predictor and the others through the random effects) helps make the LMMs comparable at a given grid point because the same sources of variation are represented. As a result, importance measures estimated using each LMM can be compared with greater confidence than

if the random effects were not used (i.e. the inclusion of the random effects reduces the differences between the LMMs at a given grid point in their qualities of fit).

With respect to interpreting the statistical model specification above, the result of using only the temporal variation P' in surface pressure as a predictor is that the effect of the variation across ensemble members in the distribution (mean) of surface pressure can be subsumed into the variation ($\psi^{(0)}$) across the members in the intercept of each statistical model. This can be seen because differences between ensemble members in mean surface pressure will have a systematic effect on the rate of precipitation in each member, and this can be captured by the random intercept, $\psi^{(0)}$, in each statistical model. We allow the variation across ensemble members in mean surface pressure to be subsumed into the variation in $\psi^{(0)}$ because only using the temporal component of surface pressure reduces the number of inputs and so simplifies the analysis. Whereas in other applications this simplification would usually be seen as a disadvantage because it restricts the detail that can be gained, in this chapter our focus is only on quantifying the variation in summer precipitation that can be accounted for by the uncertainty in each RCM parameter, and we can do this with the reduced number of inputs.

Above we described that by only using the temporal variation in surface pressure as a predictor, the effect of the variation across ensemble members in its distribution is allowed to be subsumed into the random intercept. Specifically, doing this means that any effect of the perturbations in an RCM parameter *on precipitation via surface pressure* (i.e. the effect on precipitation of the variation in surface pressure that is driven by perturbing the RCM parameter) can be captured by the importance measures of the RCM parameter, and the rest of the effect of the variation across ensemble members in the distribution of surface pressure can be subsumed into the random intercept. The result is that the variation in summer precipitation that can be accounted for by the uncertainty in each RCM parameter can be quantified, but we do not attempt to distinguish between the effect of perturbing each RCM parameter on precipitation via its effect on surface pressure and the effect of the perturbations directly on the precipitation (i.e. due to other processes than via its effect on surface pressure). This is because we are interested in how the RCM parameters affect the mean precipitation rate, and how they affect the relationship between surface pressure and precipitation, but we are not interested in the variation in mean surface pressure in this application.

Just as $\psi^{(0)}$ captures the variation in the intercept of each statistical model that is not attributable to the included RCM parameter, the second random effect term $\psi^{(1)}$ captures the variation across ensemble members in the relationship between the

surface pressure and the precipitation that is not attributable to the included RCM parameter. This is because of the interaction term between the temporal component of surface pressure and each RCM parameter captures the variation across ensemble members in the relationship between the pressure and precipitation that is attributable to the RCM parameter R. Given this, the inputs to each statistical model can be summarized as

- the perturbations across ensemble members in the included RCM parameter, R,
- the remaining variation across ensemble members in the intercept after accounting for the perturbations across ensemble members in the RCM parameter, captured by $\psi^{(0)}$,
- the remaining variation across ensemble members in the relationship between the temporal variation in surface pressure and summer precipitation after accounting for the perturbations across ensemble members in the RCM parameter, captured by $\psi^{(1)}$,
- the temporal variation in surface pressure, P',

in addition to the residual term, ϵ. Of these inputs, however, we focus on the effect of each RCM parameter R.

We use the specification above at each grid point and for each RCM parameter except for the three RCM parameters (the timescale for destruction of the CAPE, Cape, the convective updraught factor, cnv_upd, and the convective anvil shape factor, anvil) that are switched off in some ensemble members and on with different values in others. With these three RCM parameters (building on the first parameterization of the LMM shown above)

$$\mu_{ij} = \beta_0 + \beta_1 P'_{ij} + \beta_2 R_j + \beta_3 P'_{ij} R_j + \psi_j^{(0)} + \psi_j^{(1)} P'_{ij} + \beta_4 D_j + \beta_5 P'_{ij} D_j \ ,$$

where dummy variable D is introduced to indicate when the RCM parameter is switched off and R is set to zero in these ensemble members.

We use restricted maximum likelihood (REML) (Patterson and Thompson, 1971) to estimate the parameters of each LMM. To illustrate the validity of the assumptions behind the specification, diagnostic plots using raw residuals for the LMM that includes the threshold of relative humidity for cloud formation (RHCrit) at the Birmingham grid-point are shown in Figure 5.5. There is some indication that the LMM can underestimate the summer-total precipitation at fitted values of approximately 200 mm (top-left panel), and that the residual variation increases with the

fitted values (top-right panel). There is also some indication that the residual variation differs with ensemble member (middle-right panel), and that the distribution of the residuals has a heavy upper tail (bottom panels). As noted in the exploratory data analysis (Section 5.2), assuming a linear relationship between surface pressure and the precipitation, and assuming constant residual variation, are likely to be approximations to the true relationships and residual variations. However we feel that these assumptions provide a useful approximation to the true process, guarding against over-fitting and allowing the variation in precipitation that can be accounted for by the perturbations in each RCM parameter to be quantified adequately.

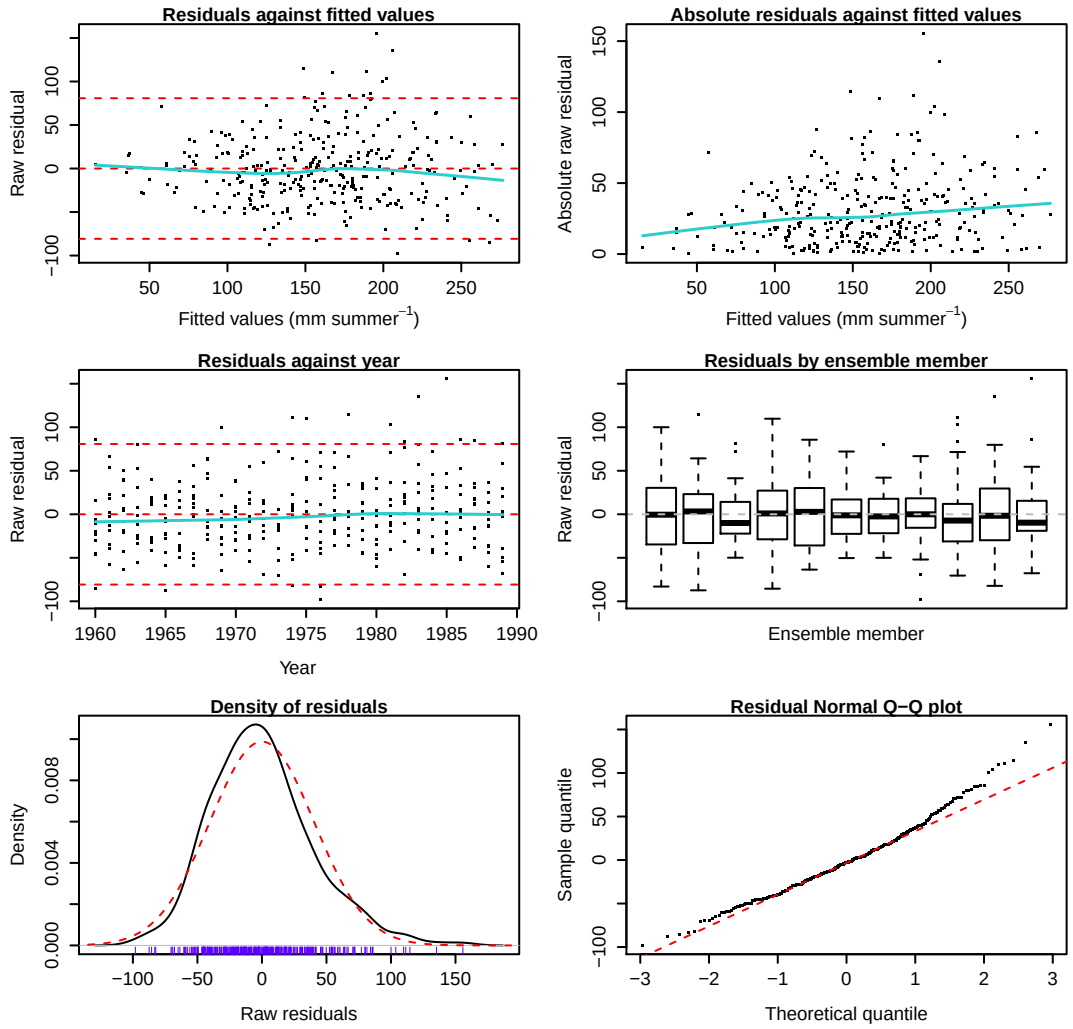

Figure 5.5: Diagnostic residual plots (using raw residuals) for the LMM that includes the RCM parameter RHCrit at the Birmingham grid-point. The dashed red lines provide an idealized guide, and the solid blue curves are local polynomial smoothers showing any trends where appropriate.

5.4 Simulation procedure

Following the methodology of Chapter 2, we use simulation to estimate the importance measures of each RCM parameter. We do this separately with each LMM at each grid point over the British Isles. The inputs (R, sometimes D, $\psi^{(0)}$, $\psi^{(1)}$, P', as well as ϵ) to each LMM have been specified in the previous section, although we focus on the effects of the RCM parameters by only estimating the importance measures of each RCM parameter. The first-order effect of an RCM parameter quantifies the variation in summer-total precipitation that can be solely accounted for by the perturbations in the RCM parameter. In addition to this, the total effect of an RCM parameter includes its interaction with the temporal variation in surface pressure, i.e. includes variation due to that across ensemble members in the relationship between surface pressure and summer precipitation that can be accounted for by the perturbations in the RCM parameter.

To estimate the importance measures of each RCM parameter, a model must be specified for all of the inputs for each LMM at each grid point. Following the recommendations of Subsection 2.2.2 and Section 2.4, we sample from the multivariate Gaussian distribution fitted to the random effects ($\psi^{(0)}$ and $\psi^{(1)}$), and assume independence between these and the temporal variation in surface pressure by sampling from the empirical distribution of the pressure (P'). We also sample from the empirical distribution of each perturbed RCM parameter. The right panel of Figure 5.2 (page 155) suggests that there is some variation across ensemble members in the shape of the distribution of P, but that the distributions are similar enough to make the assumption of independence between P' and the other inputs a useful simplification for this application. This can be seen because none of the distributions are particularly skewed, and although there appears to be some variation across the members in the variance of P, accounting for this variation across members using the methods discussed in Section 2.4 would substantially increase the complexity of the decomposition. We base each importance measure on 1500 samples of size 2000 (i.e. $R = 1500$ and $r = 2000$ using the notation of Section 2.3).

5.5 Results

5.5.1 Single grid point

Before analysing maps of the British Isles, we demonstrate the results at the Birmingham grid-point. The total variation in summer-total precipitation estimated using the LMM that includes the threshold of relative humidity for cloud formation

(RHCrit) at this grid point is 4479 mm^2 $summer^{-2}$. This is greater than the empirical variation (4215 mm^2 $summer^{-2}$) in the precipitation, but only by a ratio of 1.06. The total variation in summer-total precipitation estimated by the statistical modelling that includes each of the other RCM parameters is similar (between 4356 and 4638 mm^2 $summer^{-2}$). This suggests that the potential violations of the assumptions behind the LMMs seen in Figure 5.5 have not resulted in the total variation in the precipitation being overestimated (or underestimated) to any concerning extent.

Table 5.3 displays the importance measures of each RCM parameter at the Birmingham grid-point, as well as whether increasing (or switching on) each RCM parameter has a positive or negative association with the summer-total precipitation in the ensemble. For the three RCM parameters (Cape, conv_upd and anvil) that are switched off in some ensemble members but values vary when switched on, their effects shown in Table 5.3 are conditioned on them being switched on. The table shows that the perturbations in the ice-fall speed (VF1) (row 2) can account for the most variation (14%) in summer precipitation in the ensemble, and has a positive association with the summer-total precipitation. The little difference between the total effect (14%) and first-order effect (13%) of VF1 shows that this is largely due to the influence of the perturbations in VF1 on the mean precipitation rate, rather than on the relationship between surface pressure and precipitation. Indeed, the first-order effect of each of the RCM parameters is similar to its total effect. This suggests that the extent to which precipitation varies with the RCM parameters depends relatively little on the temporal variation in the atmospheric conditions on the seasonal scale.

Focusing on the total effects in Table 5.3, aside from the ice-fall speed (VF1), the perturbations in the number of soil levels accessible for transpiration (soillev) (row 22) can account for 12% of the total variation in summer-total precipitation in the ensemble, with a positive association. The perturbations in the cloud droplet to rain conversion thresholds (CW) (row 3) can account for 11% of the total variation in the precipitation in the ensemble, and those in the asymptotic neutral mixing length parameter (lambda) (row 18) can account for 10%, both with negative associations. By contrast, (almost) none of the variation in summer-total precipitation in the ensemble can be accounted for by the perturbations in the cloud droplet to rain conversion rate (CT) (row 4), those in the roughness length (f_rough) (row 21), the switching on or off of the surface-canopy decoupling scheme (canopy) (row 23), or the perturbations in the dynamical diffusion e-folding time (dyndiff) (row 25).

	Parameter	TE (%)	FOE (%)	Association (+/−)
Large-scale cloud				
1	RHCrit	6	6	−
2	VF1	14	13	+
3	CW	11	11	−
4	CT	0	0	−
5	eacf	5	4	−
6	rhparam	7	7	+
Convection				
7	Ent	5	5	+
8	Cape	4	4	+
9	cnv_upd	5	5	−
10	anvil	7	6	+
Radiation				
11	Icesize	4	4	−
12	ice_type	3	3	−
13	sw_absn	2	1	+
Sea ice				
14	MinSIA	5	4	−
15	oi_diff	4	4	−
16	ice_tr	5	4	+
Boundary layer and surface processes				
17	flux_g0	1	0	+
18	lambda	10	10	−
19	Charnoc	7	7	−
20	cnv_rl	7	7	+
21	f_rough	0	0	+
22	soillev	12	11	+
23	canopy	0	0	−
24	sto_res	2	1	+
Dynamics				
25	dyndiff	0	0	−
26	k	6	6	+
27	gw_lev	1	0	+

Table 5.3: Total effect (TE) and first-order effect (FOE) of each parameter (or set of parameters) of HadRM3 perturbed in the ensemble at the Birmingham grid-point, and whether increasing (or switching on) each RCM parameter has a positive or negative association (third column) with summer-total precipitation in the ensemble. The total effects and first-order effects are expressed as percentages (rounded to the nearest integer) of the estimated total variation in the precipitation.

5.5.2 Exploratory maps over the British Isles

We repeat the analysis at each grid point (separately) over the British Isles to examine the spatial pattern. The top six panels of Figure 5.6 display some exploratory maps over the study region, which can be used to characterize the region. The greatest mean (top-left panel) precipitation rates are windward of the Scottish and Welsh mountains in the west of Britain (see bottom-right panel for orography), given the prevailing south-westerly winds. The variance (top-centre panel) of the summer-total precipitation is also greatest over these areas. The lowest precipitation rates are in the south-east of the study region, particularly over the sea. The lowest proportion (top-right panel) of the summer precipitation that is large-scale (dynamical) is in southern England. The total precipitation is dominated by large-scale precipitation over the sea surrounding Scotland, and windward of the highest points of the Scottish mountains. The summer-mean surface pressure (centre-left panel) is generally greater over sea than the land, and is lowest on average up the Scottish and Welsh mountains. The variance (centre-centre panel) of the pressure depends on latitude, with the greatest variation in the north. This demonstrates that the atmospheric summer conditions are generally more settled in the south of the British Isles, at least when averaged over each summer. This, as well as the effect of the orography, is emphasized by the mean vertical velocity (see centre-right panel) over the study region, which is generally stronger in the north and the greatest ascent is windward of high ground.

Aside from the orography, the bottom panels of Figure 5.6 show some diagnostics for the total variation in the summer-total precipitation in the ensemble estimated using the LMM that includes the threshold of relative humidity for cloud formation (RHCrit). The spatial pattern of the empirical variation in the precipitation is captured by the statistical modelling, although it is estimated to be greater than the empirical over most of the study region except for the sea surrounding Scotland and west of Ireland.

Now that the study region has been characterized, and diagnostic maps have been examined, we present and describe the results at the grid-point level over the British Isles in the next subsection. We emphasize, however, that the next subsection is designed to demonstrate results that can be obtained using the methodology proposed in this chapter, rather than to necessarily be of scientific interest in themselves. This is because the small number of ensemble members relative to parameters perturbed means that we cannot assume that the results presented in this chapter will generalize to larger ensembles. We will consider in the discussion (Section 5.7) part of this chapter how the results compare to the existing literature.

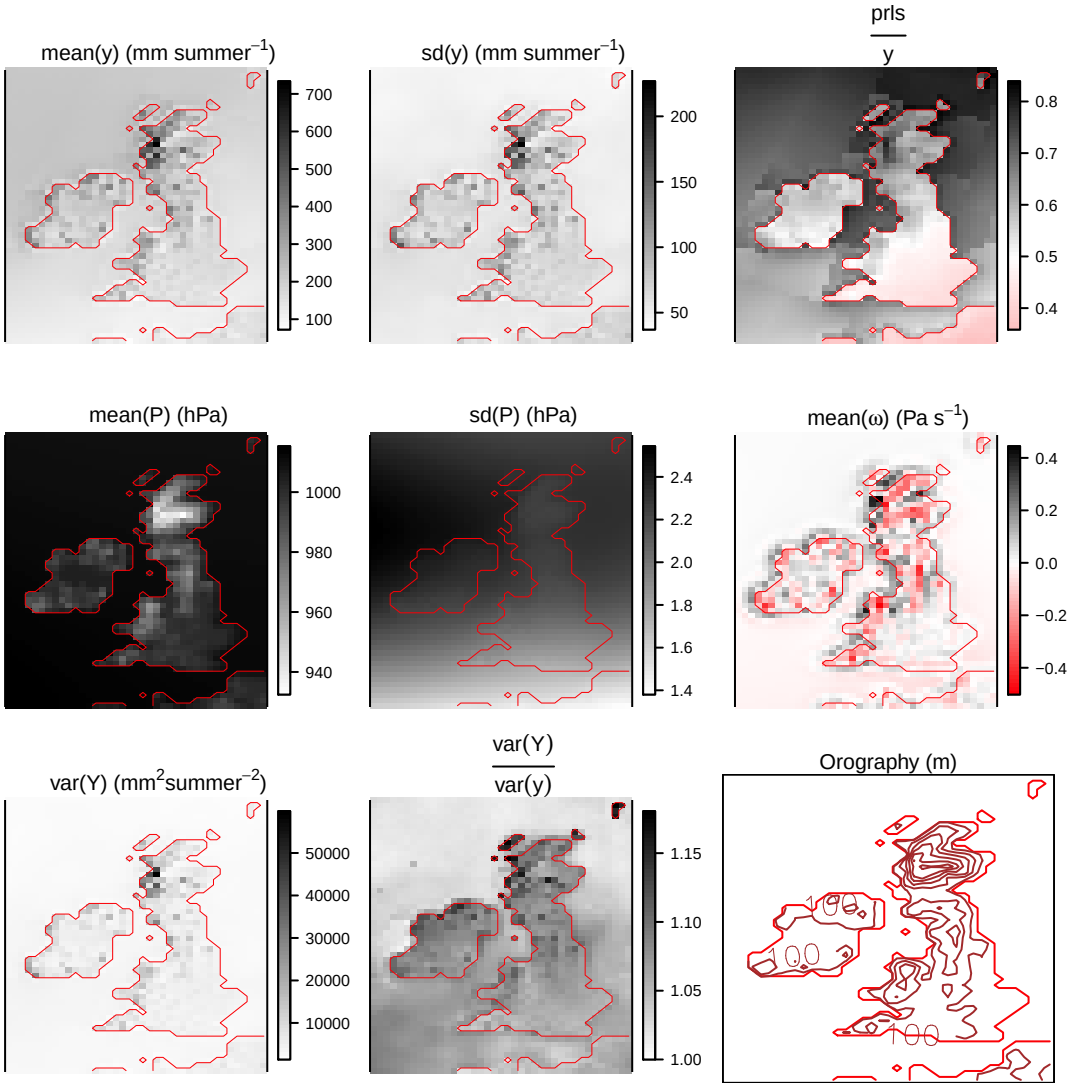

Figure 5.6: Top six panels: mean and standard deviation (sd) of the empirical summer-total precipitation (y), proportion of the precipitation that is driven by large-scale (dynamical) processes (top-right panel), mean and standard deviation of summer-mean surface pressure (P), and summer-mean ascent of air (i.e. minus the lagrangian tendency of air pressure, ω, at 850 hPa). Bottom panels: estimated total variation (var(Y)) in summer-total precipitation, and the ratio of var(Y) to the empirical variation (var(y)). Orographic contours of the British Isles are shown in the bottom-right panel. Coastline is in red.

5.5.3 Results maps over the British Isles

To see the variation in summer precipitation that can be accounted for by the perturbations in each parameter of HadRM3 in the ensemble, Figure 5.7 shows maps of the total effect of each. As with the Birmingham grid-point, the first-order effects (not shown) are very similar to the total effects over the British Isles generally. This suggests that the relationship between air pressure and the precipitation is relatively robust to the perturbations in the RCM parameters on this scale.

Figure 5.7 highlights that there can be considerable spatial variation in the total variation in summer-total precipitation accounted for by the perturbations in some

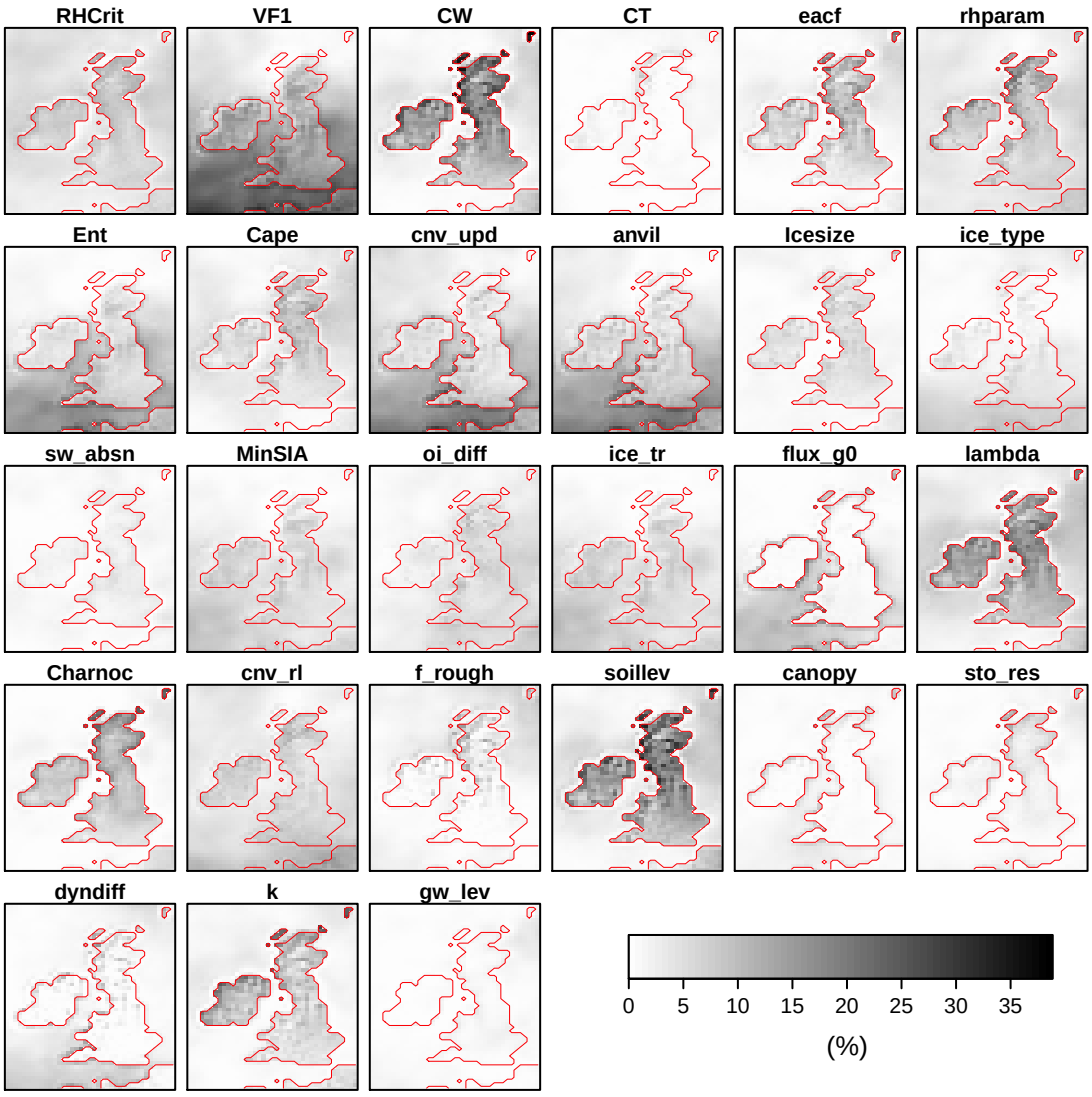

Figure 5.7: Total effects of each parameter perturbed in HadRM3 in the ensemble at the grid-point level over the British Isles. The effects are expressed as percentages of the estimated total variation in summer-total precipitation. Coastline is in red.

of the RCM parameters. At the Birmingham grid-point the perturbations in the ice-fall speed (VF1) could account for the most variation (14%) in the precipitation in the ensemble. Over the British Isles generally, VF1 can account for up to 32% of the variation in the summer-total precipitation at the grid-point level. This is highest over the sea around southern Britain, but is also relatively high over most of the land, including up the Scottish and Welsh mountains, but not along some coastlines such as the west-facing coastlines in the north. These coastlines are the areas of the British Isles where the most air is advected from the sea to over the land. The result is that these areas tend to have high rates of precipitation, and large-scale precipitation is more prominent than convective. This suggests that immediately after air is advected from the sea and up over the land, the simulation of precipitation in the ensemble is robust to the uncertainty in VF1, but not further inland.

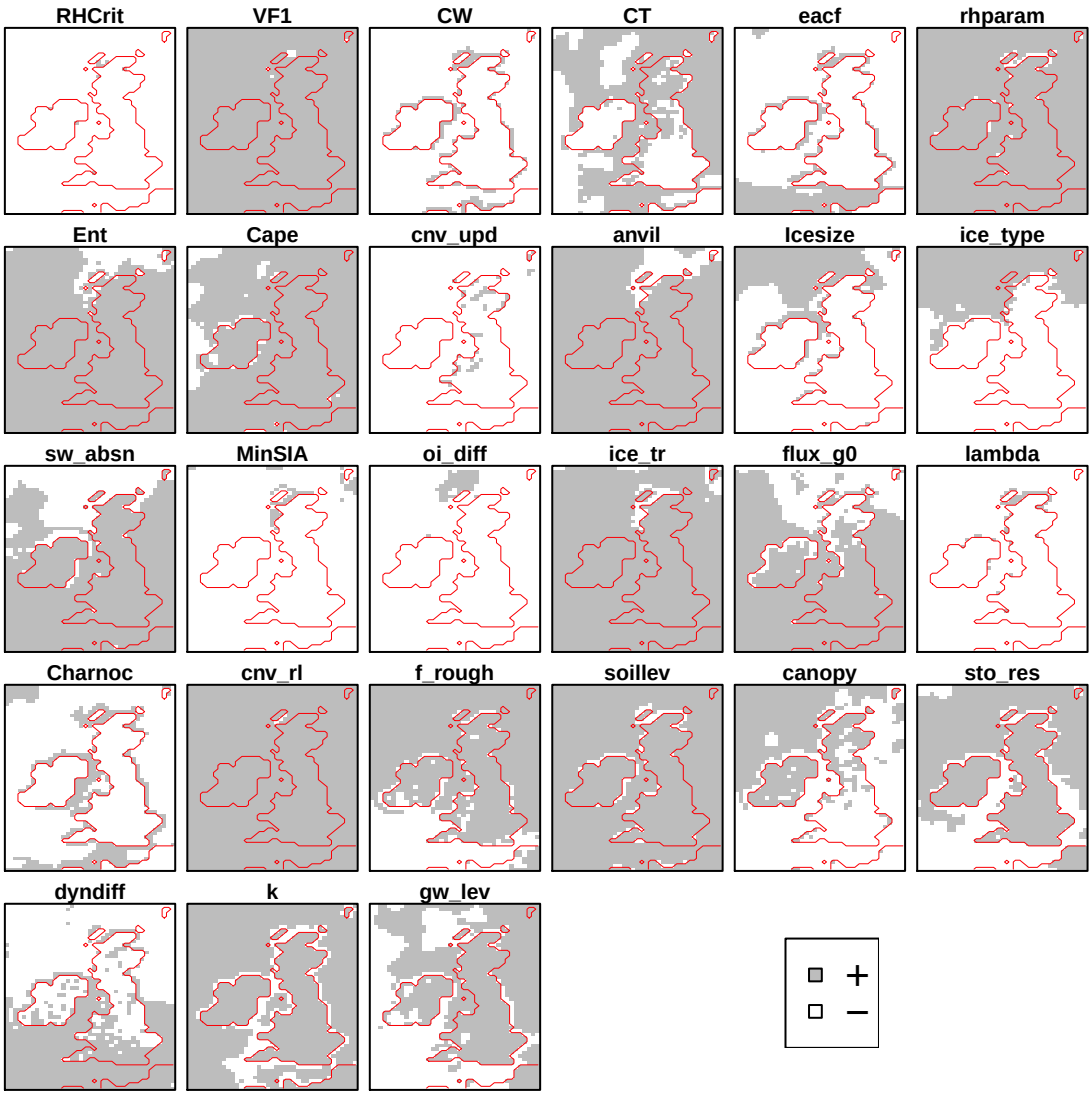

Figure 5.8: Whether each parameter perturbed in HadRM3 has a positive or negative association with summer-total precipitation in the ensemble at the grid-point level over the British Isles. Coastline is in red.

The ice-fall speed can affect the period of time for which air remains moist (recall from Subsection 1.4.2), with lower values of VF1 leading to the air remaining moister for longer. Therefore the lack of variation in the precipitation that can be accounted for by perturbing VF1 along west-facing coastlines in the north of the British Isles could be because air is often forced to rise abruptly by the land and so moisture often precipitates out regardless of the impact of VF1 on its efficiency. Whether increasing (or switching on) each RCM parameter has a positive or negative association with summer-total precipitation in the ensemble over the study region can be seen in Figure 5.8, and this shows that VF1 is positively associated with summer-total precipitation over almost all of the British Isles.

The variation in precipitation that can be accounted for by the perturbations in the

entrainment coefficient (Ent), convective updraught factor (cnv_upd) and convective anvil shape factor (anvil) all show a similar spatial pattern to that of the ice-fall speed (VF1). This is perhaps not surprising because the perturbations in these four RCM parameters share strong positive correlation in the ensemble. However, Ent, con_upd and anvil influence convection in the RCM, and indeed their perturbations can account for more variation in the precipitation in the south of the study region where convective precipitation is more prominent than in the north. Increasing or switching on Ent and anvil generally have a positive association with summer-total precipitation in the ensemble, but there is some suggestion that increasing conv_upd when it is switched on leads to a reduction in the precipitation. The perturbations in Ent, cnv_upd and anvil cannot, however, account for as much variation as those in VF1 over most of the study region.

Unlike with the ice-fall speed (VF1), the greatest proportion of the total variation in summer-total precipitation in the ensemble that can be accounted for by the perturbations in the cloud droplet to rain conversion thresholds (CW) (up to 39%) is over the land of the British Isles. The greatest proportion accounted for is in the north and along some of the west-facing coastlines, as well as over the islands north of the Scottish mainland. This is also the case with the number of soil levels accessible for transpiration (soillev) (up to 34%) and the asymptotic neutral mixing length parameter (lambda) (up to 26%). There is strong correlation between the perturbations in these three RCM parameters in the ensemble, with CW and lambda generally associated negatively with the precipitation and soillev associated positively. Given the spatial pattern of their total effects, this suggests that when CW and soillev are low and lambda is high, a large amount of precipitation is simulated by HadRM3 when air is advected from sea to over land. Low CW means that the moisture that is held in the atmosphere by clouds falls as precipitation more readily (i.e. it requires less collision and coalescence, recall from Section 1.3), and so when air is forced to rise due to advection from the sea it will precipitate more readily near the coast instead of further inland. The significance of the perturbations in soillev and lambda is perhaps less clear than with CW when air is advected from sea over land. Indeed, the high proportion of variation in precipitation that their perturbations can account for may simply be due to their correlation with those in CW in the ensemble, particularly because soillev is perhaps unlikely to be important near coastlines where moisture is readily available.

The perturbations in several parameters of HadRM3 contribute little to the variation in summer-total precipitation in the ensemble over the study region. These RCM parameters are the cloud droplet to rain conversion rate (CT), shortwave water vapour continuum absorption on/off (sw_absn), surface-canopy decoupling scheme

on/off (canopy), stomatal conductance response to ΔCO_2 on/off (sto_res), and the gravity wave drag start level (gw_lev). The perturbations in the boundary layer flux profile parameter (flux_g0) and the dynamical diffusion e-folding time (dyndiff) also contribute little to the variation in the precipitation in the ensemble over the land of the British Isles, but can account for more of the variation over the sea to the south-west of Britain. The perturbations in the roughness length (f_rough) account for little of the variation in the precipitation over much of the British Isles, but can account for up to 15% over Scotland and the sea immediately around it, particularly along the west coast of Scotland. This is the area that is the most affected by orography due to the mountains in Scotland and its relatively high latitude.

The perturbations in the cloud fractions at grid-box saturation (eacf), ice particle size (Icesize), Charnock constant (Charnoc) and the wave constants (k) can each account for more of the variation in precipitation in the ensemble over the land of the British Isles, particularly in the north, than over the sea. The correlations between the perturbations in these RCM parameters are not strong, except for that (0.78) between eacf and Charnoc. Increasing eacf, Icesize and Charnoc have a negative association with the precipitation in the ensemble over most of the study region, whereas increasing k has a negative association with precipitation just off many of the coastlines of the British Isles but positive elsewhere.

In addition to eacf, Icesize, Charnoc and k, the perturbations in the threshold of relative humidity for cloud formation (RHCrit), flow-dependent RH_{crit} scheme on/off (rhparam), timescale for destruction of CAPE (Cape), and the free convective roughness length over sea (cnv_rl) can each generally account for more of the variation in the precipitation in the ensemble over the land of the British Isles than the sea. This contrast between land and sea is often more pronounced particularly in the north, but is less pronounced than with eacf, Icesize, Charnoc and k. Increasing RHCrit is negatively associated with summer-total precipitation in the ensemble over all of the British Isles, whereas switching rhparam on is generally positively associated, as is increasing Cape and cov_rl.

The perturbations in the non-spherical ice particles (ice_type), sea ice albedo (MinSIA), ocean-ice diffusion coefficient (oi_diff) and the sea ice minimum temperature (ice_tr) can generally each account for more of the variation in precipitation in the ensemble in the south of the study region than in the north. The correlation between the perturbations in MinSIA and ice_tr is almost perfect (-0.99), but the other correlations between these RCM parameters are not particularly strong. The association between the precipitation and increasing ice_type in the ensemble is negative apart from over some of the sea to the north of Scotland and Ireland. MinSIA

and oi_diff are generally negatively associated with the precipitation, but increasing ice_tr is generally associated with more summer precipitation. With relatively few ensemble members, however, we cannot determine whether these are causal effects or simply due to correlations with other RCM parameters.

5.6 Summary

The perturbations in the parameters of HadRM3 that can account for the greatest proportion of variation in summer-total precipitation in our perturbed physics ensemble depends on the area of the British Isles. Summers in the north of the British Isles are less settled (wetter and windier) than in the south. Due to this and the prevailing south-westerly winds, west-facing coastlines in the north experience the most advection of air from sea to over the land. Along these coastlines, the perturbations in the cloud droplet to rain conversion thresholds (CW) can account for the greatest proportion (up to 39%) of variation in summer-total precipitation, followed by the perturbations in the number of soil levels accessible for transpiration (soillev) (up to 34%) and those in the asymptotic neutral mixing length parameter (lambda) (up to 26%). The perturbations in these three RCM parameters are, however, strongly correlated, and so we cannot disentangle their individual effects in this ensemble. In particular, it is perhaps unlikely that the number of soil levels accessible for transpiration (soillev) is important near these coastlines because moisture is readily available from the sea.

The perturbations in the ice-fall speed (VF1) account for little variation in the precipitation in the ensemble along the west-facing northern coastlines of the British Isles. Despite this, the perturbations in VF1 can account for the greatest proportion of the variation in the precipitation compared to the other RCM parameters over much of the land of southern Britain (up to 26%) and the surrounding sea (up to 32%). This spatial pattern is also seen with the entrainment coefficient (Ent), convective updraught factor (cnv_upd) and the convective anvil shape factor (anvil). The perturbations in these three RCM parameters cannot account for as much of the variation in the precipitation in the ensemble as those in VF1, however, and they share strong correlations with each other and with the perturbations in VF1. Despite the correlations, Ent, cnv_upd and anvil influence convection in the RCM, and so are likely to contribute to the variation in summer-total precipitation in the south of the British Isles, where convection is more prominent than in the north.

Whereas increasing the cloud droplet to rain conversion thresholds (CW) is generally negatively associated with summer-total precipitation in the ensemble, increasing the ice-fall speed (VF1) is positively associated with the precipitation over the

British Isles. Unlike with these RCM parameters, the perturbations in many of the parameters of HadRM3 contribute little to the variation in the precipitation over the land of the British Isles and much of the sea. These RCM parameters are the cloud droplet to rain conversion rate (CT), shortwave water vapour continuum absorption on/off (sw_absn), surface-canopy decoupling scheme on/off (canopy), stomatal conductance response to ΔCO_2 on/off (sto_res), gravity wave drag start level (gw_lev), and to a lesser extent the boundary layer flux profile parameter (flux_g0) and the dynamical diffusion e-folding time (dyndiff).

5.7 Discussion

By using statistical modelling and estimating importance measures we have shown how to quantify the extent to which uncertainty in each parameter of an RCM contributes to the variation in its simulation of summer precipitation. In doing so we have found that perturbing the RCM parameters affects the represented atmospheric conditions (and therefore the precipitation) and/or additional precipitation bias, but we have not attempted to distinguish between these effects in the analysis. We have identified the ice-fall speed (VF1) and the cloud droplet to rain conversion threshold (CW) as the RCM parameters whose perturbations can account for the most variation in summer-total precipitation in a perturbed physics ensemble of HadRM3, depending on spatial area of the British Isles. However this ensemble only has eleven members and 27 of the RCM parameters are independently perturbed. The result is that there are strong correlations between the perturbations in the RCM parameters in the ensemble. More ensemble members would be required to disentangle the individual effect of the uncertainty in each RCM parameter, and so we cannot assume that our results generalize to other or larger ensembles. Indeed, with more ensemble members we could have included all of the perturbed RCM parameters in the same statistical model and considered interactions between them. With a large enough ensemble, including all of the RCM parameters in the same statistical model would mean that the effect of the uncertainty in each could be corrected for by the statistical model, and the orthogonal individual effect of the uncertainty in each RCM parameter could be quantified. Note, however, that the importance of an RCM parameter in an ensemble depends on the magnitude of its perturbations in the ensemble, and so the results would only generalize to other ensembles in which the magnitudes of the perturbations were similar to those in the analysed ensemble.

Although we cannot assume that our results generalize to other ensembles due to the correlations between the perturbations in the RCM parameters, some of the results support previous studies from the literature. Finding that the perturbations

in the ice-fall speed (VF1) can account for the most variation in the summer precipitation in our ensemble over much of the south of the British Isles supports the literature that VF1 is an important parameter in climate models with respect to their simulation of precipitation (see Subsection 1.4.2 and references therein). As also outlined in Subsection 1.4.2, the entrainment coefficient is often found to influence convective precipitation in climate models. In our ensemble we found that the perturbations in the entrainment coefficient can account for up to 13% of the variation in the summer-total precipitation in southern Britain, and up to 26% over the surrounding sea. Its perturbations cannot, however, account for as much of the variation in the precipitation as those in VF1. Over the north of the land of the British Isles, we found that the cloud droplet to rain conversion thresholds (CW) can account for the most variation in summer-total precipitation in the ensemble, and this threshold (which differs over land and sea) is one of the climate model parameters that we identified in Subsection 1.4.2 as relevant to precipitation. Of the other climate model parameters identified in Subsection 1.4.2, we found that the perturbations in the number of soil levels accessible for transpiration (soillev) can account for the most variation in summer-total precipitation in the ensemble over much of the land of the British Isles, but this may be due to the strong correlation between its perturbations and those of CW.

Future work is to use the methodology set out in this chapter with a larger ensemble to orthogonally decompose the total variation in summer-total precipitation into that due to the uncertainty in each RCM parameter individually, as well as any interactions between them. By doing this the significance of the uncertainty in each parameter of the RCM could be determined more generally, and this could be compared to other climate models to generalize the findings further. Additionally, the extent to which the uncertainty in each climate model parameter affects its simulation of precipitation could be investigated on different temporal scales, e.g. daily, because analysing summer totals sums over the variation that occurs on shorter temporal scales.

6 Extremes

Aim: To extend the methodology of Chapter 2 to focus on extreme precipitation, and to use the extended methodology to quantify the extent to which the simulation of extreme winter precipitation events is robust to the uncertainty in the physical parameters of climate models.

6.1 Introduction

Extreme climatic events, including precipitation extremes, can be devastating to society and the natural world. This has led to studies focusing on only extreme climatic events (e.g. see Karl et al., 1997; Easterling et al., 2000). Therefore in this chapter we extend the methodology of Chapter 2 to focus on extreme precipitation, and use the extended methodology to quantify the extent to which the simulation of extreme winter precipitation events varies across the members of a perturbed physics ensemble over the British Isles.

The need to analyse extreme precipitation has led to the use of generalized extreme value (GEV) or generalized Pareto (GP) statistical models fitted to observations and climate model output of extreme events (e.g. Yee and Stephenson, 2007; Fowler and Ekström, 2009; Fowler et al., 2010; Min et al., 2011). Indeed, the use of extreme value modelling (Fisher and Tippett, 1928) to analyse extreme events is supported by statistical theory (Coles, 2001; Katz et al., 2002). One way of quantifying the extent to which the representation of driving physical processes of extreme precipitation varies across climate models is, therefore, to relate predictor variables to the extreme events using an extreme value statistical mixed model and apply the methodology of Chapter 2. To this end, however, if a generalized linear mixed model (GLMM) that is fitted to the whole distribution of the precipitation captures the extreme events well, then fitting a separate statistical model to only the extreme events is not necessary. As we show in this chapter, this is because the methodology of Chapter 2 can be extended to focus on the extremes.

The methodology of Chapter 2 uses a GLMM fitted to the precipitation that is simulated on a given spatial and temporal scale by an ensemble of climate models to decompose the total variation in the precipitation. This includes the extreme precipitation events, of course, and so the same GLMM can be used to focus on these extreme events. Furthermore, if the GLMM captures the extremes well, then an advantage to using it to condition on the extremes is that the conditional decomposition could be compared to the unconditional decomposition without the need to

fit separate statistical models. Comparing such decompositions could be informative in itself, but could also help understand why the representation of a physical process is sensitive to the differences between the climate models, and when the sensitivity is relevant. For example the sensitivity of the simulation of precipitation may be found to be due to disagreement between the climate models in their simulation of extreme events, which may only be relevant in certain atmospheric conditions.

After outlining the extension to the methodology to focus on extremes in Section 6.2, we demonstrate the extended methodology in Section 6.3 using the numerical example of Chapter 2. The application to extreme winter precipitation events over the British Isles is then in Section 6.4. Note that, unlike in previous chapters, we exclude subscripts on most conditional expectations and conditional variances for brevity in this chapter (otherwise these subscripts can be long when focusing on extremes).

6.2 Focusing on extremes: methodology

Extreme value modelling with the generalized Pareto family of distributions involves analysing peaks (excesses) over a specified high threshold. The methodology of Chapter 2 can be extended to analyse extremes by conditioning the decomposition on the occurrence of an extreme precipitation event, i.e. on the amount of precipitation (Y) being over the specified threshold (t), $Y > t$. Recall from Section 2.2 that we use the law of total variance to decompose the total variation in precipitation, which allows the importance measures of the inputs to the GLMM to be estimated. For example the decomposition that includes the first-order effect (FOE($\boldsymbol{X'}$), $\text{var}_{\boldsymbol{X'}}[E_{\mu}(\mu \mid \boldsymbol{X'})]$) of the temporal components ($\boldsymbol{X'}$) of the predictors is shown by Equation (2.11) on page 53. Recall that μ is the expectation of the precipitation, such that $\mu = E(Y \mid \mu)$. Now let μ_t be the conditional expectation of the precipitation, such that $\mu_t = E(Y \mid \mu\,;\, Y > t)$. Focusing on the precipitation extremes by conditioning each term of the decomposition on $Y > t$ gives (excluding subscripts)

$$
\begin{aligned}
&\text{var}(Y \mid Y > t) \\
&\quad = \text{var}[E(Y \mid \mu\,;\, Y > t) \mid Y > t] + E[\text{var}(Y \mid \mu\,;\, Y > t) \mid Y > t] \\
&\quad = \text{var}(\mu_t \mid Y > t) + E[\text{var}(Y \mid \mu\,;\, Y > t) \mid Y > t] \\
&\quad = \text{var}[E(\mu_t \mid \boldsymbol{X'}\,;\, Y > t) \mid Y > t] + E[\text{var}(\mu_t \mid \boldsymbol{X'}\,;\, Y > t) \mid Y > t] \\
&\qquad + E[\text{var}(Y \mid \mu\,;\, Y > t) \mid Y > t]\,.
\end{aligned}
$$

Recall from the methodology outlined in Chapter 2 that the fitted expectation μ of the precipitation does not depend on the shape (e.g. variance) of the fitted distribution. Moreover we showed that this means that the estimation of FOE(μ) and the residual variation (TE(ϵ)) is unaffected by any dependence between the fitted expectation μ and the residual term ϵ in the unconditional (with respect to extremes) case, since μ and ϵ are uncorrelated. This is not necessarily the case when conditioning on an extreme precipitation event, however, because the conditional expectation does depend on the shape of the fitted distribution in the general case. We highlight this later (in the numerical example) for the Gaussian family, for example, but this can be seen more generally because, given μ and a high threshold t, a wider distribution will often result in more of the distribution being above the threshold.

Given the above, unlike in the unconditional general decomposition outlined in Chapter 2, dependence between the residual term ϵ and the other inputs could affect the estimation of the initial decomposition of the conditional variation in Y into $\text{var}(\mu_t \mid Y > t)$ and the remaining variation shown above. This can be seen because if expectation μ is less than the threshold t, then ϵ must be positive for an extreme to occur. Despite this, $\text{var}(\mu_t \mid Y > t)$ is the reduction in the conditional variation in precipitation Y expected by fixing μ constant, and, similarly to the unconditional case, we can define this as the explained variation in Y given $Y > t$. Furthermore $E[\text{var}(Y \mid \mu\,; Y > t) \mid Y > t]$ is the reduction in the conditional variation in Y from averaging over ϵ, and we can define this as the unexplained (residual) variation.

Having defined $\text{var}(\mu_t \mid Y > t)$ as the explained variation in the conditional decomposition of precipitation Y, we can then proceed in following the structure of the unconditional decomposition of Chapter 2 by decomposing this explained variation. Therefore, similar to the unconditional case, we define total effects in a conditional decomposition to be on μ_t to be consistent with the interpretation of $\text{var}(\mu_t \mid Y > t)$ as the conditional variation explained by the GLMM. In Chapter 2, decomposing the explained variation ensured that any dependence between the residual term ϵ and the other inputs did not affect the estimation of the total effects in the unconditional case. When conditioning on extremes, however, the effect of such dependence on the estimation of importance measures (including first-order effects) cannot be avoided since the dependence can affect the estimation of $\text{var}(\mu_t \mid Y > t)$, as seen above. Therefore we define all importance measures to be those obtained from decomposing the explained variation in the conditional case. We can, however, interpret the importance measures in the conditional case in a similar way as in the unconditional case when there is dependence between inputs (recall Subsections 1.10.2, 2.2.2 and 2.3.2 for discussion on importance measures in the presence of dependent in-

puts), i.e. by taking their mathematical definitions as informative but recognising that dependence could affect their estimation. The difference between the conditional and unconditional cases is that the dependence could be with ϵ, in addition to inputs to μ, in the conditional case.

With respect to estimating the conditional terms, it follows from the mathematics above that two adjustments must be made to the importance measures when focusing on the precipitation extremes. These are 1) the expectation μ must be converted into the conditional expectation μ_t given μ and $Y > t$, and 2) the conditional distribution of each input given $Y > t$ must be used. In other words, focusing on extremes means that the tail expectation above the threshold must be used, as well as the distributions of the inputs conditioned on an extreme precipitation event (which will not necessarily be the tails of their own distributions).

In Chapter 2 we outlined how simulation could be used to estimate the importance measures. With respect to using simulation to estimate the conditional importance measures, the first adjustment ((1) above) required to focus on extremes means finding the relationship between the unconditional expectation and the tail expectation (above the threshold) for the family of distributions used by the GLMM. To see how this can be done, suppose that a two-parameter (μ and λ) family of distributions (e.g. Gaussian or Gamma) is used to model precipitation. In this case the corresponding λ could be sampled for each sample of μ, thereby defining a distribution for the precipitation. If the relationship is closed form then it can be used to convert each sampled value of μ into μ_t. If the relationship is not closed form then simulation can be used for each conversion. Sampling from the distribution defined by each sample of μ, and computing the mean of the samples greater than the threshold, would provide the μ_t that corresponds to the sampled μ. This procedure would increase the sampling error of the simulation, as well as the computing expense, but is rarely necessary because Landsman and Valdez (2005) provide closed form relationships of the Gaussian, Gamma, Inverse Gaussian, Poisson, Binomial and Negative Binomial families of distributions.

The unconditional simulation procedure requires specifying a model for the inputs and sampling from it (recall from Section 2.4). The samples are then used to compute the importance measures by taking means and variances as appropriate (Subsection 1.10.3). The second adjustment ((2) above) to the simulation procedure to focus on extremes is to sample from the distributions of the inputs that lead to extreme precipitation. This does not, however, require specifying a different model for the inputs to that specified when analysing the unconditional precipitation. Instead, not only can the unconditional model be used for the inputs, the same samples can

even be used. This is because the samples only need to be weighted when computing the means and variances required for each importance measure. As detailed in the next paragraphs, the weights are the estimated probability that each sample of inputs will lead to extreme precipitation.

In the unconditional simulation procedure, each sample of inputs is used to compute a value for the expected precipitation, μ. As stated previously, if the family of distributions that is used by the GLMM is defined by more than one parameter (e.g. the Gaussian and Gamma families are defined by two parameters each), and the other parameters of the family vary with the inputs, then these parameters can also be computed from each sample. Along with μ, these parameters define a probability distribution for the precipitation Y for each sample of inputs. The estimated probability that each sample will lead to extreme precipitation can be computed from each distribution, thereby providing the weight for each sample that is required to focus on the extreme precipitation events. Mathematically, these weights are

$$\Pr(Y > t \mid \mu, \lambda)$$

for families of distributions defined by two parameters, its expectation (μ) and dispersion parameter (λ). Weighting the samples in this way respects the conditional distributions of the inputs, as discussed below.

To see why weighting the samples by their estimated probability of leading to an extreme Y respects the conditional distributions of the inputs, consider the equivalence of this method to the alternative (but perhaps more intuitive) method of simulating a value of Y from each sample of μ and λ, and discarding all samples that do not lead to an extreme Y. We do not recommend this alternative method because discarding samples is computationally wasteful and could lead to taking means and variances based on small sample sizes when computing importance measures, whereas weighting each sample by the probability of an extreme Y is mathematically equivalent but avoids these issues. This equivalence is clear because sampling Y for each μ and λ and only retaining each sample such that $Y > t$ means retaining each sample of μ and λ with its probability of leading to an extreme Y before computing the importance measures, and this is equivalent to weighting all of the samples by these probabilities when performing the computation.

To illustrate how to use simulation to estimate importance measures when focusing on extremes, the following outlines the simulation procedure for estimating the first-order effect ($\mathrm{var}[E(\mu_t \mid X\,; Y > t) \mid Y > t]$) of an input X, where the family of distributions used by the GLMM has two parameters (expectation μ and dispersion λ), both of which vary with X. Note that whereas computing dispersion λ was

not required in the methodology of Chapter 2, λ is required here to, along with μ, define a full probability distribution for each sample to evaluate μ_t and compute the estimated probability ($\Pr(Y > t \mid \mu, \lambda)$) of leading to extreme Y.

- Randomly generate a value for X from its specified (unconditional) distribution.
- Randomly generate r values for each of the other inputs from their specified (unconditional with respect to extremes) joint distribution, given the value for X.
- Compute the corresponding r values for μ and λ.
- Compute the corresponding r values for μ_t.
- For each sample of μ and λ, and using the cumulative distribution function of the family of distributions assumed for Y by the GLMM, compute the corresponding probability (w_1) of an extreme Y occurring, such that $w_1 = \Pr(Y > t \mid \mu, \lambda)$.
- Compute the mean (w_2) of the r probabilities (w_1) of extreme Y.
- Using w_1 as weights, estimate $E(\mu_t \mid X\,; Y > t)$ by the weighted mean (W) of the r values for μ_t, i.e.
$$W = \frac{1}{r} \sum_{i=1}^{r} \mu_{t_i} \frac{w_{1i}}{w_2} \,,$$
where i indexes the sample of μ_t and w_1.
- Repeat the previous steps R times, obtaining R values for W and w_2, each at different X.
- Using w_2 as weights, estimate $\text{var}[E(\mu_t \mid X\,; Y > t) \mid Y > t]$ by the weighted variance of the R values for W. This weighted variance is
$$\frac{1}{\sum_{j=1}^{R} w_{2j} - \frac{\sum_{j=1}^{R} w_{2j}^2}{\sum_{j=1}^{R} w_{2j}}} \sum_{j=1}^{R} w_{2j} \Big(W_j - \overline{W} \Big)^2 \,,$$
where j indexes the sample of W and w_2, and $\overline{W} = \frac{1}{R} \sum_{j=1}^{R} W_j \frac{w_{2j}}{\frac{1}{R} \sum_{j=1}^{R} w_{2j}}$ is the weighted, by w_2, mean of W (see page 266 of Galassi et al., 2013, who note that this estimator reduces to the usual unweighted sample variance estimator when the weights are equal and non-zero).

With respect to interpreting the conditional decomposition, the dependence structure of the inputs to μ will generally be different when conditioned on extreme

precipitation events to when unconditioned. This means that if independence was assumed between these inputs for an unconditional decomposition, dependence between the inputs will often be induced when extending the decomposition to focus on extremes. To see this, consider a scenario where two (equally important) independent inputs ($X^{(1)}$ and $X^{(2)}$) determine a variable (Y), and suppose that both have a positive relationship with Y. In this scenario, conditioning on the extremes of Y will mean that $X^{(1)}$ and $X^{(2)}$ will become negatively related because sampling a low $X^{(1)}$ will mean that $X^{(2)}$ must be high, and vice versa. Such dependence will often mean that an orthogonal decomposition is not obtainable when conditioning on $Y > t$, although non-orthognal decompositions generally remain informative (Subsections 1.10.2, 2.2.2 and 2.3.2).

The dependence between inputs due to conditioning on $Y > t$ cannot necessarily be avoided by using a different method such as fitting a statistical model to only extremes. This can be seen because conditional inputs $(X^{(1)} \mid Y > t)$ and $(X^{(2)} \mid Y > t)$ will always be negatively related in the scenario above. Moreover, the dependence structure given a precipitation extreme may be weaker than in the unconditional case, and the decomposition of the extremes will sometimes be simpler than the unconditional decomposition because variation in some inputs may be important in the unconditional case but unimportant in the extremes. Therefore although the dependence structure of the inputs will generally be different when conditioned on extreme precipitation to when unconditioned, the conditional decomposition will not necessarily be more complicated to interpret.

Focusing on extremes using the methodology proposed here can increase the number of inputs that interact with each other. If the relationship between the unconditional expectation (μ) of the precipitation and its conditional expectation (μ_t) is non-linear over the range of the conditional inputs, then this will mean that all inputs interact. Although such interactions would not be present if a statistical model without a non-linear link function is fitted to only the extremes, all inputs interact in any statistical model that uses a non-linear link function to relate μ to the response variable, and so the presence of interactions between all inputs will not necessarily complicate a conditional decomposition.

The conditional dependence structure between the inputs can be seen from the samples taken for the simulation procedure. Samples can be plotted against each other and their correlations can be examined. In addition to using importance measures, the interactions between the inputs can be examined using the samples. This can be done in the same way as for the unconditional decomposition, i.e. by plotting $\text{var}(\mu_t \mid \cdot\,; Y > t)$ against the different inputs and examining any relationships (recall

Figure 2.8, page 69, in Section 2.3).

Despite the dependence structure and often the number of interactions differing between the conditional and unconditional cases, there are advantages to fitting a GLMM to the whole distribution of precipitation and using it to focus on the extremes. Firstly, more of the data can be used to estimate the relationships between the predictors and the precipitation. Secondly, it is generally desirable to check the sensitivity of a statistical model to the choice of threshold, and indeed the analysis may be of interest in itself using many different thresholds. This would require fitting, evaluating, and sampling from a different statistical model for each different threshold if using statistical models fitted to only extremes. When using a single GLMM fitted to the whole distribution of precipitation, however, it is trivial to estimate conditional importance measures using different thresholds. The same samples can be used from the same model, because (as seen above) only the weightings and conditional expectations need to be altered when estimating the importance measures. This is also computationally cheap relative to fitting and sampling from multiple statistical models.

6.3 Numerical example

As an illustrative example of focusing on extreme events, we extend the numerical example of Chapter 2 (recall Section 2.3) by analysing events above a high fixed threshold (t) and comparing the importance measures to their unconditional equivalents, i.e. to when analysing the whole distribution of the response variable (Y). Recall that the statistical model is

$$
\begin{aligned}
Y_{ij} &= 3 + 6X_{ij}^{(1)} + 2X_{ij}^{(2)} + \psi_j^{(0)} + \psi_j^{(1)}X_{ij}^{(1)} + \psi_j^{(2)}X_{ij}^{(2)} + \epsilon_{ij} \\
X_{ij}^{(1)} &= \gamma_j^{(1)} + X_{ij}^{\prime(1)} \qquad \gamma_j^{(1)} \sim N(1, 36) \\
\psi_j^{(0)} &\sim N(0, 25) \qquad X_{ij}^{\prime(1)} \sim N(0, 9) \\
\psi_j^{(1)} &\sim N(0, 16) \qquad X_{ij}^{(2)} \sim N(1, 4) \\
\psi_j^{(2)} &\sim N(0, 9) \qquad \epsilon_{ij} \sim N(0, 100) \ ,
\end{aligned}
$$

where i indexes time and j indexes the member of an ensemble of climate models. The residual term is ϵ and we denote the expectation of Y given the inputs as μ, $\mu = E(Y \mid \mu)$.

6.3.1 Exploratory data analysis

Recall that we demonstrated how the sources of variation in Y could be visualized

in Subsection 2.3.1 by randomly generating a data set from the known model and plotting the data. We now use this data set to visualize how the sources of variation differ between the conditional and unconditional cases. We define an extreme Y as greater than the 95th percentile of its empirical sample in the data set, such that $t = 57.5$. The top panels of Figure 6.1 show the unconditional relationships between Y and each predictor. Recall that the dominant sources of variation in Y are the variation across ensemble members in the distribution (mean) of $X^{(1)}$, the temporal variation in $X^{(1)}$, and the variation across ensemble members in the relationship between $X^{(1)}$ and Y. The variation in $X^{(2)}$, and the variation across ensemble members in its relationship with Y, are relatively unimportant. See the first column of Table 6.1 for a reminder of some sources of variation in the unconditional case. In the table we denote all of the interactions involving (e.g.) $X^{(1)}$ as $\text{int}(X^{(1)}, \cdot)$. Note that in the unconditional decomposition $X^{(1)}$ and $\psi^{(1)}$ only interact with each other, so $\text{int}(X^{(1)}, \cdot)$ equals $\text{int}(\psi^{(1)}, \cdot)$, and this is also the case with $X^{(2)}$ and $\psi^{(2)}$.

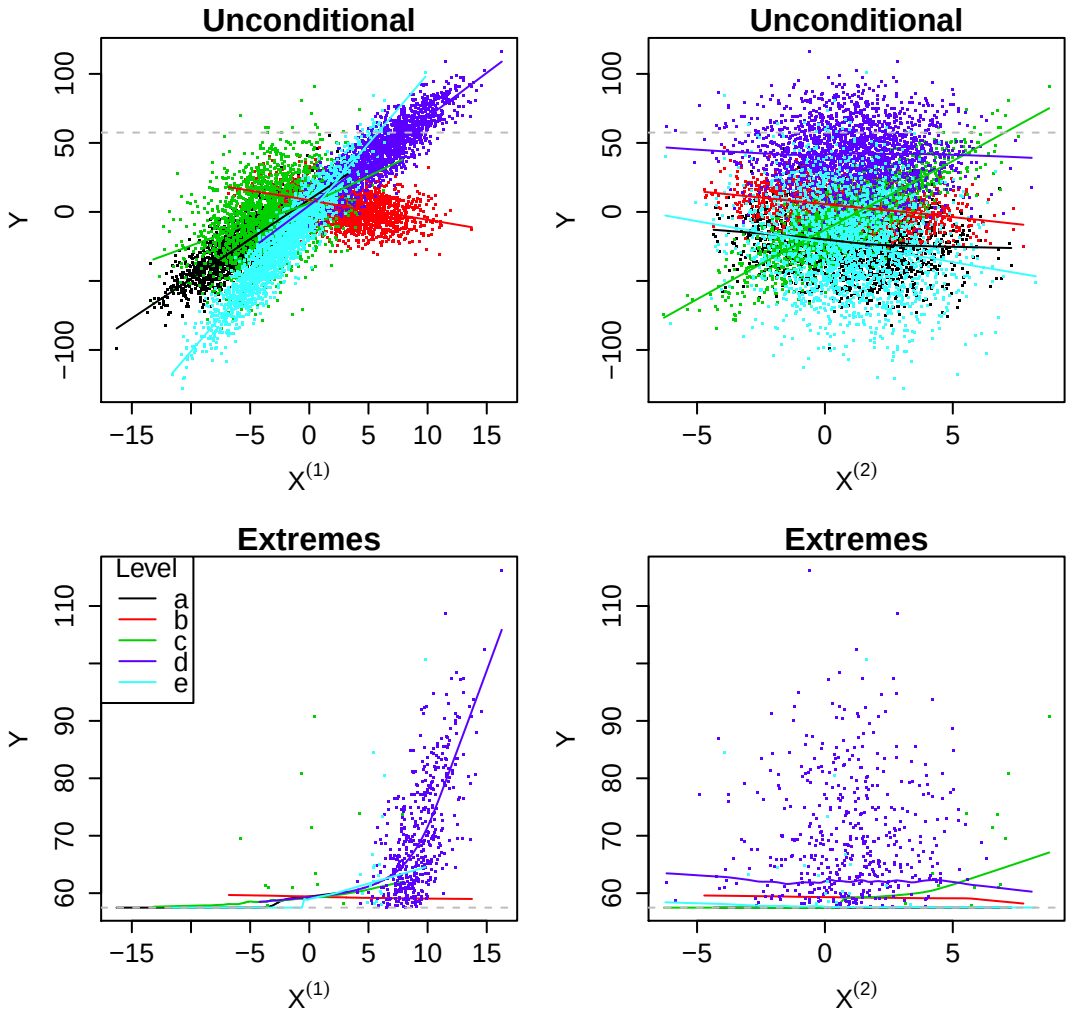

Figure 6.1: Response Y against predictors ($X^{(1)}$ and $X^{(2)}$) in the unconditional case (top panels) and when focusing on the extremes (bottom panels) of Y, coloured by level of the random factor (ensemble member). Curves are lowess smoother approximations of $\mu = E(Y \mid \mu)$ for the unconditional case and $\mu_t = E(Y \mid \mu\,;\, Y > t)$ for the extremes.

	Unconditional	Extremes
$\text{var}(Y)$	2542	1296
Residual variation	4	6
$\text{FOE}(X'^{(1)})$	13	3
$\text{FOE}(X^{(2)})$	1	0
$\text{FOE}(\gamma^{(1)})$	51	19
$\text{FOE}(\psi^{(0)}, \psi^{(1)}, \psi^{(2)})$	2	18
$\text{int}(X^{(1)}, \cdot)$	28	39
$\text{int}(X^{(2)}, \cdot)$	1	3
$\text{int}(\psi^{(0)}, \cdot)$	0	1
$\text{int}(\psi^{(1)}, \cdot)$	28	47
$\text{int}(\psi^{(2)}, \cdot)$	1	2

Table 6.1: Global sensitivity analysis results for the numerical example, showing first-order effects (FOE), interactions (int) and the residual variation, all expressed as percentages of the total variation in Y. With the interactions, (e.g.) $\text{int}(X^{(1)}, \cdot)$ denotes all interactions between $X^{(1)}$ and the other inputs, and so some are non-orthogonal. First column shows the unconditional case; second column focuses on an extreme response variable Y. All values are rounded to zero decimal places.

The bottom panels of Figure 6.1 focus on the extremes of Y. Extreme Y generally occur at high $X^{(1)}$ and when its relationship with Y is strong and positive. Given an extreme Y, the dominant source of variation appears to be the interaction between $X^{(1)}$ and its relationship with Y. The extent to which Y depends on the ensemble member at $X^{(1)} < 7$ seems small, whereas at (e.g.) $X^{(1)} = 15$ the variation in Y is dominated by its relationship with $X^{(1)}$. The strong positive relationship between $X^{(1)}$ and Y in level d leads to Y expected to be approximately 105, but Y is expected to be less than 70 in the other levels. As in the unconditional case, relatively little of the variation in Y appears to be accounted for by the variation in $X^{(2)}$. Therefore, Figure 6.1 suggests that the main differences between the unconditional and conditional cases appear to be that the variation in $X^{(1)}$ alone accounts for less of the variation in Y given an extreme Y, but more of the variation in Y is due to the interaction between $X^{(1)}$ and its relationship with Y in the extremes.

6.3.2 Simulation procedure

We use simulation to estimate the importance measures for extreme Y by sampling from the true model and following the methodology outlined in Section 6.2 to decompose the conditional variation in Y, given $Y > t = 57.5$. In doing this, each sample of the expectation (μ) of Y must be converted into its corresponding conditional

expectation (μ_t). We use

$$\mu_t = \mu + \frac{\varphi(z)}{1 - \Phi(z)}\sigma \tag{6.1}$$

for this, where φ is the standard Gaussian probability density function, Φ is the standard Gaussian distribution function, σ is the standard deviation of the distribution, and $z = (t - \mu)/\sigma$ is the number of standard deviations the threshold (t) is from the mean of the distribution. This relationship is a property of the Gaussian family of distributions (see Landsman and Valdez, 2003).

The black curves in Figure 6.2 show unconditional kernel density estimates of the values sampled for each input. The red curves in the figure show the conditional ($Y > t$) distributions, achieved by retaining each sample with probability equal to its probability of leading to an extreme Y. The figure highlights that extreme values of Y generally occur as a result of high $X^{(1)}$ (due to high $\gamma^{(1)}$) and/or a strong positive relationship ($\psi^{(1)}$) between $X^{(1)}$ and Y.

The top-left panel of Figure 6.3 shows that there is non-linearity between the expectation (μ) of Y and its corresponding conditional expectation (μ_t). When μ is above the threshold it (approximately) equals μ_t, but μ_t is close to the threshold when μ is below it. The bottom-left panel demonstrates that μ is below the threshold in a substantial proportion (12%) of the samples that lead to extreme Y (due to the residual term). The non-linearity between μ and μ_t means that all of the inputs interact with each other in the decomposition of the extremes (just as when a non-linear link function is used in a statistical model to relate μ to Y).

Despite the additional interactions between the inputs when focusing on the extremes of Y, Figure 6.4 shows that the interactions involving the relationship ($\psi^{(1)}$) between $X^{(1)}$ and Y are dominated by its interaction with $X^{(1)}$. Similar to Figure 2.8 (page 69), Figure 6.4 shows this by displaying the total variation in the conditional expectation (μ_t) of Y due to the variation in $\psi^{(1)}$, plotted against samples of the other inputs. Recall that strong relationships indicate strong interactions because it shows that the variation in μ_t (and therefore Y) due to the variation in $\psi^{(1)}$ depends on the other input. In the unconditional case $\psi^{(1)}$ only interacts with $X^{(1)}$, and although $\psi^{(1)}$ interacts with each of the other inputs given an extreme Y, its interaction with $X^{(1)}$ remains dominant in the conditional case. Similarly, the interactions involving $X^{(1)}$ are dominated (not shown) by that with its relationship ($\psi^{(1)}$) with Y.

With respect to the dependence between inputs, the panels on the right of Figure 6.3

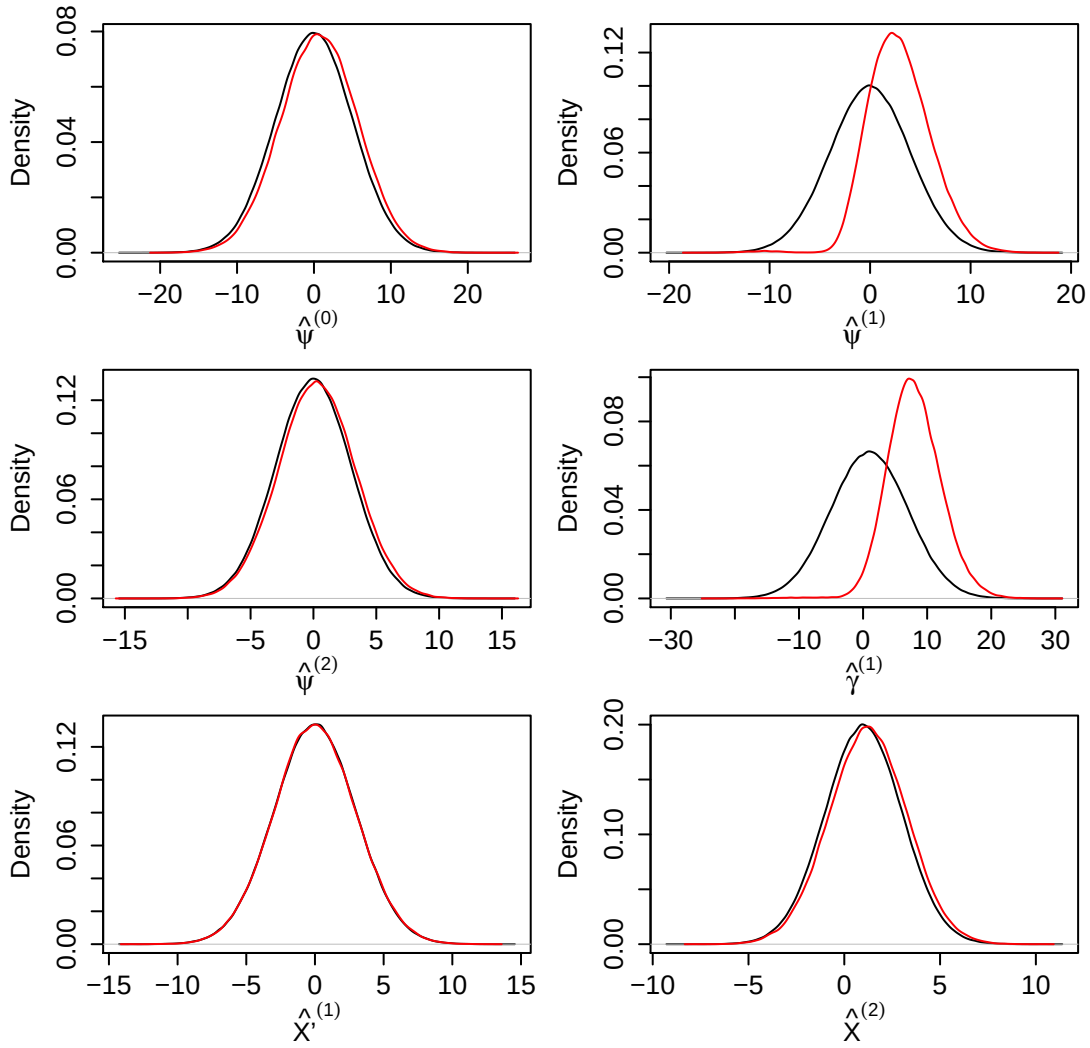

Figure 6.2: Kernel density estimates for the inputs of the numerical example. Red curves are conditional on an extreme ($Y > t$) value of Y and black curves are unconditional.

use samples taken of $X^{(1)}$ and $X^{(2)}$ to illustrate how the dependence can change. Whereas $X^{(1)}$ and $X^{(2)}$ are independent in the unconditional case (top-right panel), they are negatively related given an extreme Y. This is because although the variation in $X^{(2)}$ is relatively unimportant, high values usually contribute a little to the occurrence of extreme Y, which tend to occur at high $X^{(1)}$. With regard to the other inputs, the correlation matrix between the conditional samples is

$$\operatorname{cor}\begin{pmatrix} X^{(1)} \\ X^{(2)} \\ \psi^{(0)} \\ \psi^{(1)} \\ \psi^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & -0.06 & -0.05 & -0.12 & -0.05 \\ -0.06 & 1 & 0.00 & -0.08 & 0.15 \\ -0.05 & 0.00 & 1 & -0.07 & 0.00 \\ -0.12 & -0.08 & -0.07 & 1 & -0.07 \\ -0.05 & 0.15 & 0.00 & -0.07 & 1 \end{pmatrix}.$$

The greatest correlations between the inputs in the conditional decomposition are 0.15 between $X^{(2)}$ and $\psi^{(2)}$, and -0.12 between $X^{(1)}$ and $\psi^{(1)}$. This shows that only weak correlations are present between the conditional inputs.

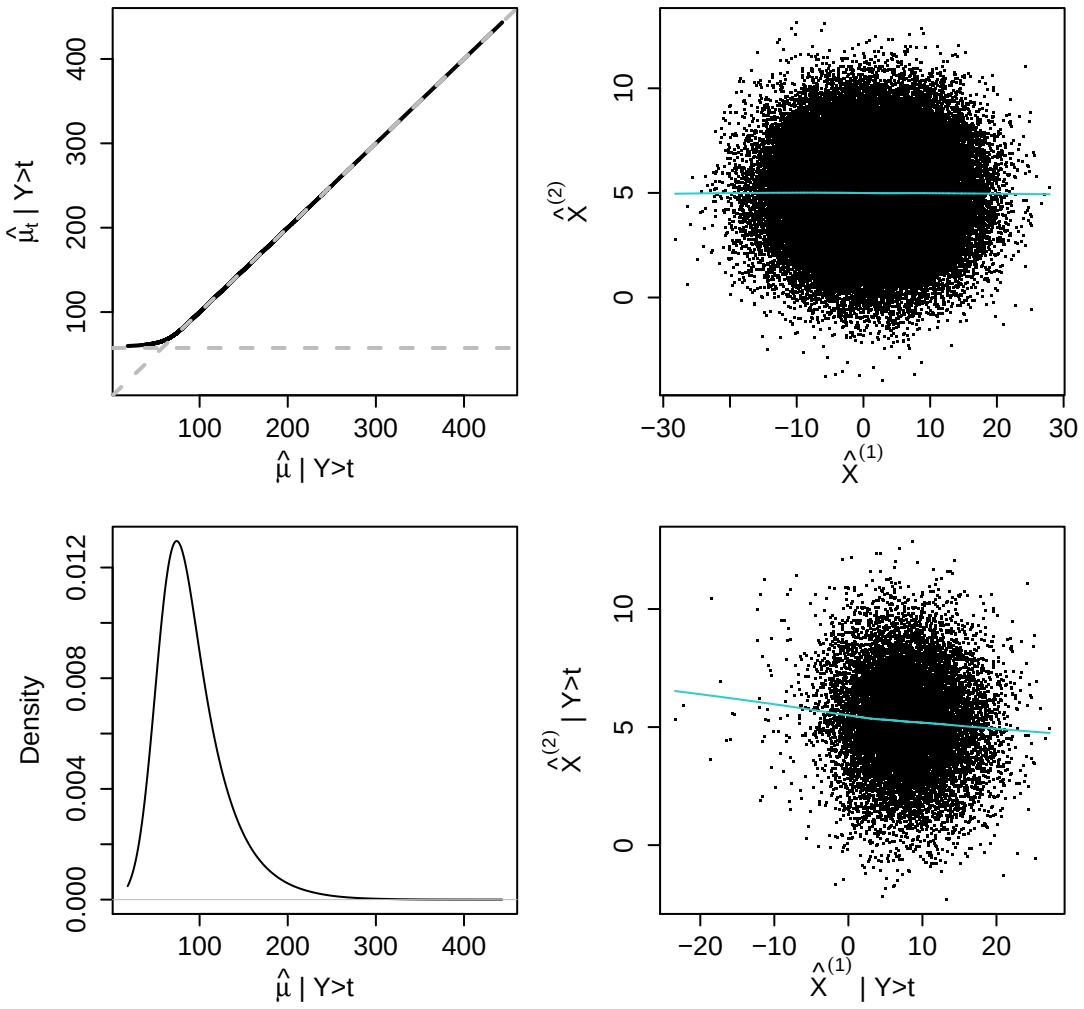

Figure 6.3: Top-left panel: sampled values of μ that led to an extreme Y against their corresponding μ_t; horizontal grey line is at the threshold (t) and the diagonal is for $\mu_t = \mu$. Bottom-left panel: kernel density estimate of μ_t from top-left panel. Right panels: sampled values of $X^{(1)}$ against those of $X^{(2)}$ for all Y (top right) and only those that led to extreme Y (bottom right).

6.3.3 Results

The second column of Table 6.1 shows the importance measures when focusing on only extreme Y. The conditional variation (1296) in Y is lower than the unconditional (2542). The proportion of the variation (FOE($\psi^{(0)}, \psi^{(1)}, \psi^{(2)}$)) across ensemble members in the bias of Y remaining after averaging over the total variation in the predictors is greater (18%) when focusing on the extremes of Y than in the unconditional case (2%). As shown by the first-order effects of $X'^{(1)}$ and $\gamma^{(1)}$, the proportion (22%) of the conditional variation in Y solely due to the variation in $X^{(1)}$ is less than in the unconditional case (64%). As in the unconditional case, little of the conditional variation in Y can be accounted for by the variation in $X^{(2)}$.

Aside from FOE($\psi^{(0)}, \psi^{(1)}, \psi^{(2)}$), increases in the importance measures from the

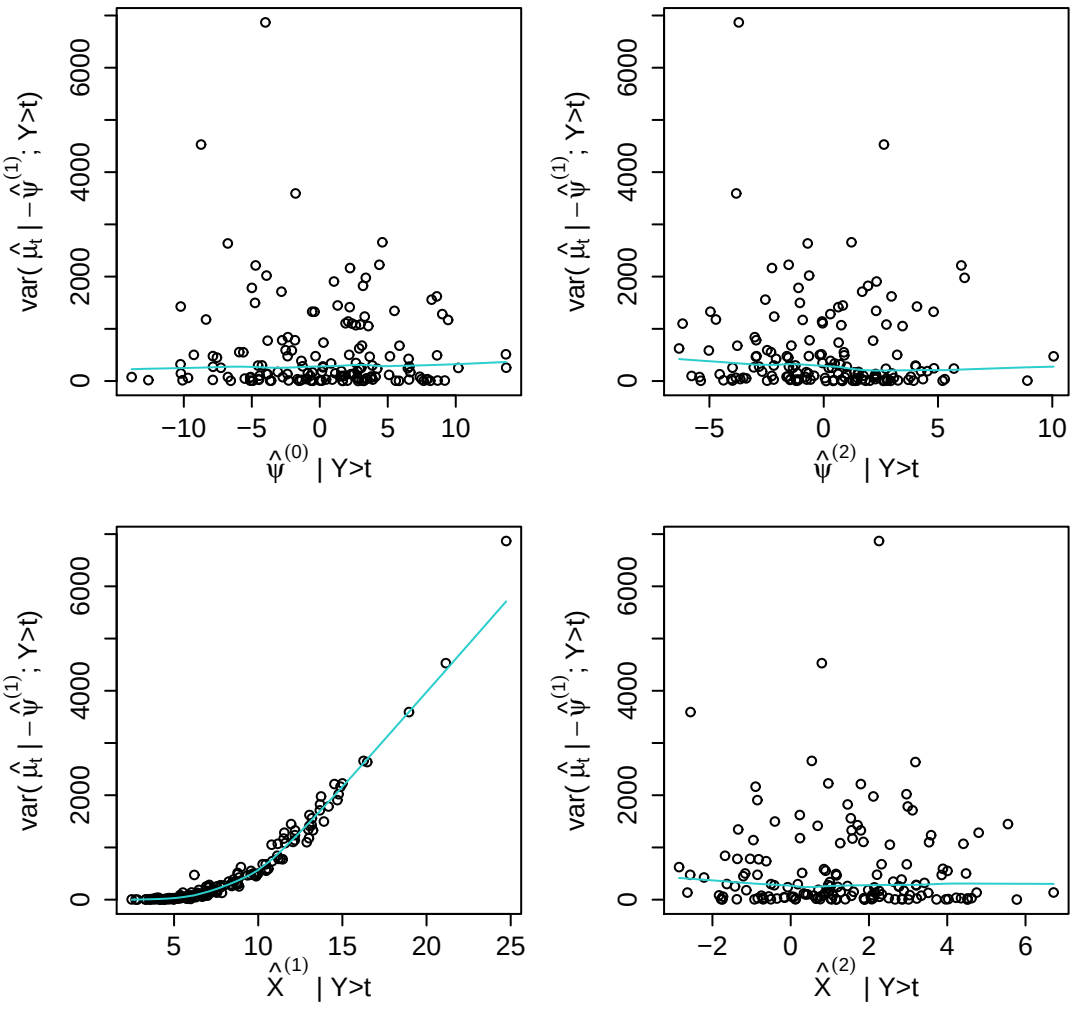

Figure 6.4: Estimated total variation ($\text{var}(\mu \mid -\psi^{(1)}\,;\, Y > t)$) in the conditional expectation μ_t of Y due to the variation in $\psi^{(1)}$, plotted against simulated values of the other inputs. Curves are local polynomial smoothers.

unconditional case to the conditional case are the interactions ($\text{int}(X^{(1)}, \cdot)$) involving $X^{(1)}$ and those ($\text{int}(\psi^{(1)}, \cdot)$) involving its relationship ($\psi^{(1)}$) with Y. These are estimated to be 39% and 47% respectively. This is because the interaction between $X^{(1)}$ and its relationship with Y is particularly important (as seen previously by Figure 6.4) in determining the magnitude of the extremes of Y. The interactions ($\text{int}(\psi^{(0)}, \cdot)$) involving the variation in the intercept, and those ($\text{int}(\psi^{(2)}, \cdot)$) involving the relationship between $X^{(2)}$ and Y, are relatively unimportant given an extreme Y.

Overall, the importance measures have quantified what was suggested by the exploratory plots of Figure 6.1, i.e. that the magnitude of an extreme Y is determined to a large extent by the variation in $X^{(1)}$ when its relationship with Y is strong and positive. Indeed, extreme events are often a result of high $X^{(1)}$ and a strong relationship with Y. Given an extreme Y, we estimate that approximately 40% of

the variation in Y is due to the interaction between $X^{(1)}$ and its relationship with Y.

6.3.4 Evaluation

The variation in $X^{(1)}$ alone accounted for much of the variation in Y when analysing the whole distribution of Y. When focusing on extreme Y, however, the variation in $X^{(1)}$ alone accounts for proportionately less of the conditional variation in Y. In contrast, more of the conditional variation in Y is due to the interaction between $X^{(1)}$ and its relationship with Y than in the unconditional case. Moreover, although the interactions between other inputs contribute little to the conditional variation in Y, there are more interactions present than in the unconditional case. The greater importance of many interactions between inputs is perhaps not surprising because for the magnitude of an extreme to be high, multiple inputs are often required to be high. In the case of this numerical example, the most extreme events occur when both $X^{(1)}$ is high and its relationship with Y is strong and positive.

A fixed threshold, above which an extreme is defined, has been used in this numerical example to illustrate the extension to the methodology to focus on extreme events. However the extended methodology is not restricted to a fixed threshold. The threshold could vary with one or more of the inputs, for example, and the only adjustment required to the simulation procedure would be to compute and use the threshold given each set of sampled inputs, rather than using the same threshold for each set of samples. Varying the threshold in this way could, for example, be useful when analysing precipitation in different seasons and an annual cycle is present in the data.

The extension to the methodology has been presented as a way of focusing on the extreme part of the distribution of precipitation. However the extension to the methodology can be used to focus on any part of the distribution of precipitation. Indeed, it can be used to focus on a part of the precipitation defined by one (or more) of the inputs, including its expectation (μ). For example, if a GLMM is designed to predict future events, then it may be of interest to decompose the variation in the precipitation (Y) given that an extreme event (above threshold t) is predicted, i.e. to decompose $\text{var}(Y \mid \mu > t)$.

6.4 Extreme winter precipitation

Extreme precipitation can cause devastating flooding in the mid-latitudes, including the British Isles (see Hall et al., 2005). Aside from localized convective extremes in summer, strong winter storms from the Atlantic Ocean can result in days of heavy precipitation over large areas of the British Isles (e.g. see Hand et al., 2004; Marsh, 2004; Lavers et al., 2011). The multiplicative effect of the vertical ascent of air and specific humidity can be seen as the main driver of precipitation (e.g. see Wilson and Toumi, 2005, and Sapiano et al., 2006) because the storms force moist air to rise, either due to weather fronts or orography of the land, such that the moisture condenses and falls usually as large-scale precipitation (recall Section 1.3). Therefore if climate models are to be used to understand these driving physical processes of extreme winter precipitation, and make useful projections of extreme events in the future, then the extent to which extreme winter precipitation events are simulated robustly across climate models is of interest.

We quantify the extent to which extreme large-scale precipitation events are simulated robustly by relating the multiplicative effect of ascent and specific humidity to pentad (five-day mean) precipitation averaged over the British Isles, and apply the methodology of Section 6.2. Given the above, analysing winter pentad precipitation averaged over the British Isles is designed to capture distinct (extreme) large-scale precipitation events. In addition to this, the analysis of winter precipitation events allows a comparison between the results of this section and those of Chapter 4, in which the robustness of convective precipitation events was analysed. Such a comparison is reserved for Chapter 7.

By using the multiplicative effect of ascent and specific humidity as the predictor of large-scale precipitation in this application, the aim is to account for much of the variation in the precipitation with the statistical model. This would allow a formal quantification of the extent to which large-scale precipitation events are simulated robustly across the ensemble members over the British Isles. An advantage of such a quantification with the methodology of Section 6.2 is that the quantification can be estimated using many different thresholds and viewed against the thresholds. This allows the extent to which the robustness depends on the severity of the events to be investigated, for example it could be that extreme events are simulated less robustly because they are caused by extreme atmospheric conditions which could exacerbate the effects of perturbing the RCM parameters. Such an investigation could help reveal why and in what conditions large-scale precipitation varies across climate models.

6.4.1 Data

To quantify the extent to which the simulation of extreme winter precipitation events is robust to uncertainty in the physical parameters of climate models, we use the same perturbed physics ensemble of the regional climate model (RCM) HadRM3 as in Chapters 4 and 5 (see Section 4.2 and references therein), and the same years of study (1960-1989). The original data are gridded across the British Isles. Therefore to analyse winter pentad precipitation averaged over the British Isles we first take five-day means for the precipitation at each grid point for the extended winter months (October to March), before then averaging over the land of the British Isles.

As in Chapter 4, we use predictors at 850 hPa to summarize the free atmosphere without being overly sensitive to surface conditions and parameterizations. To capture the multiplicative effect of vertical ascent and specific humidity at each grid point, we multiply the vertical ascent (captured by minus the Lagrangian tendency of air pressure, which measures the rate of change of air pressure following motion of the air) (ω) by specific humidity (q) and use this as a single predictor (ωq). The advantage of multiplying ascent and specific humidity into a single predictor is that we capture the multiplicative effect in greater detail by multiplying the variables on the daily scale, before aggregating up to five-day means to relate to the pentad precipitation. When multiplying daily-mean ascent and daily-mean specific humidity, we censor the daily-mean ascent at each grid point from below at zero (i.e. ωq is set to zero when $\omega < 0$). This is because positive ascent is required for precipitation, and stronger ascent will generally drive more precipitation. By contrast, descent does not drive precipitation, and so the strength of descent is of little importance. Moreover large-scale precipitation can be driven by ascent at adjacent grid points (recall the top-left and centre-right panels of Figure 5.6, page 169, from Chapter 5), meaning that precipitation can occur at a grid point where the air descends. Therefore censoring ascent at zero at the grid-point level before spatially averaging ensures that positive ascent is not cancelled out by descent elsewhere. Furthermore, unlike when spatially averaging the precipitation, we include the sea immediately surrounding the land of the British Isles when spatially averaging the predictor (see Figure 6.5 for detail of the study region).

6.4.2 Exploratory data analysis

Plotting the data to guide the statistical modelling, Figure 6.6 shows the relationship between pentad precipitation and moist ascent (i.e. ascent multiplied by specific humidity) averaged over the British Isles. As expected, there is a positive relationship in each ensemble member. There is little precipitation when there is little moist as-

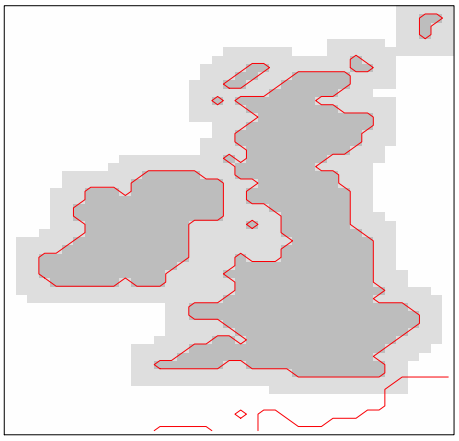

Figure 6.5: Study region (darker grey) over which precipitation is averaged (i.e. the land of the British Isles), and the study region (darker grey and lighter grey) over which the predictor is averaged (i.e. the land of the British Isles and the immediately surrounding sea). Coastline is in red.

cent, whereas pentad precipitation greater than 9 mm day^{-1} generally occurs at high moist ascent. The relationship appears to be close to linear, although there is some non-linearity in some ensemble members. The variation in precipitation around the relationship is lowest at low precipitation and moist ascent within each ensemble member, and generally appears symmetrical around each curve. There seems little variation across ensemble members in the relationship, although perhaps the most is at high moist ascent.

Figure 6.7 highlights that there is little variation across ensemble members in the distributions of the precipitation and moist ascent. Heavy precipitation events can occur in any ensemble member, but there are two extreme outlying events above 12 mm day^{-1}. There are no such extreme outliers in the moist ascent.

6.4.3 Statistical model specification

To quantify the sources of variation in winter pentad precipitation over the British Isles, we require a statistical model fitted to the data. The relationships between moist ascent and precipitation shown in Figure 6.6 appear to be approximately linear. However, capturing the extremes well is particularly important in this application, and the non-linearity seen in the relationship in some ensemble members (e.g. the top-right panel of Figure 6.6) suggests that assuming linearity would overestimate the strength of the relationship at strong moist ascent. Therefore it is desirable to capture the non-linearity in the relationship using the statistical model in this application.

Instead of assuming linearity, we use natural cubic spline bases (see Subsection 4.4.1)

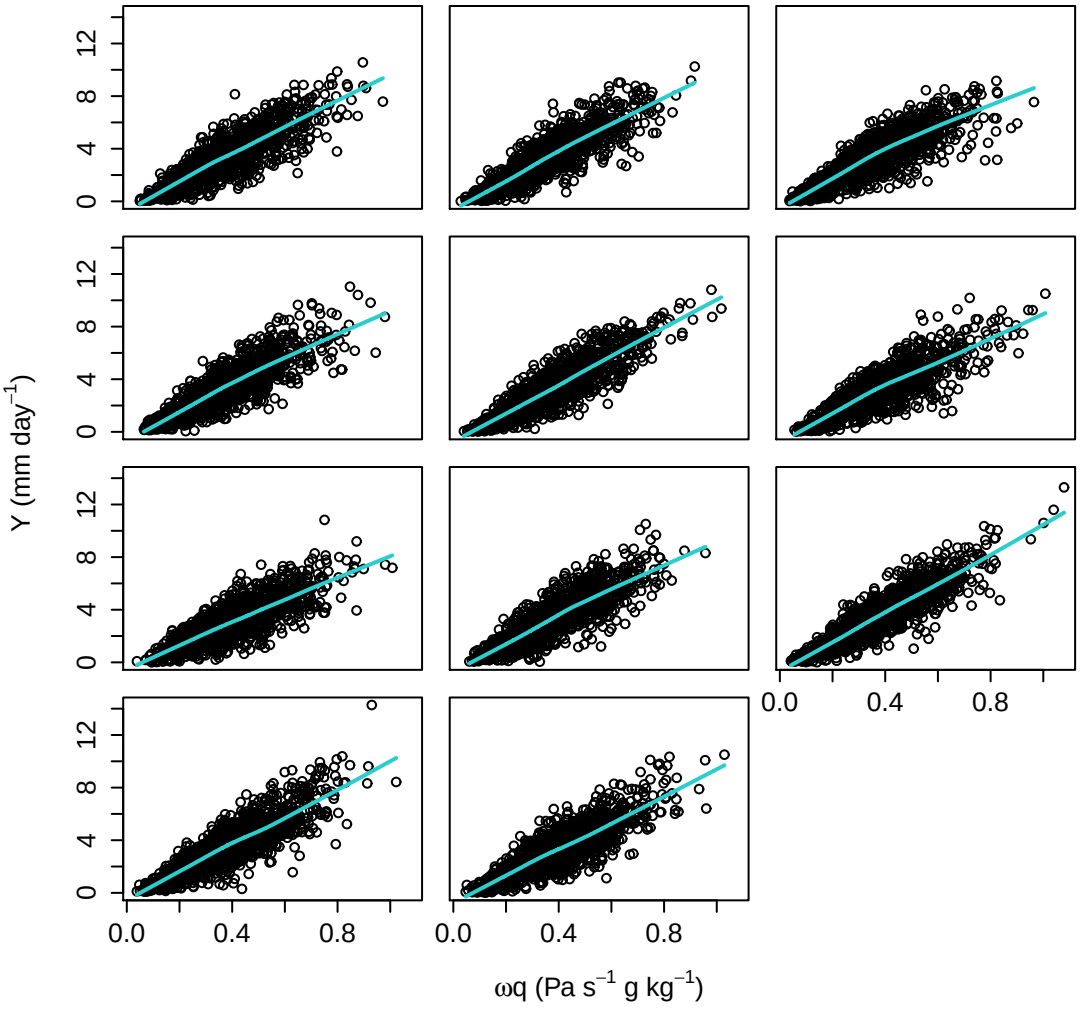

Figure 6.6: Pentad precipitation (Y) against five-day mean moist ascent (ωq) (ascent multiplied by specific humidity) in each ensemble member averaged over the British Isles. Curves are local polynomial smoothers. The control run is top left.

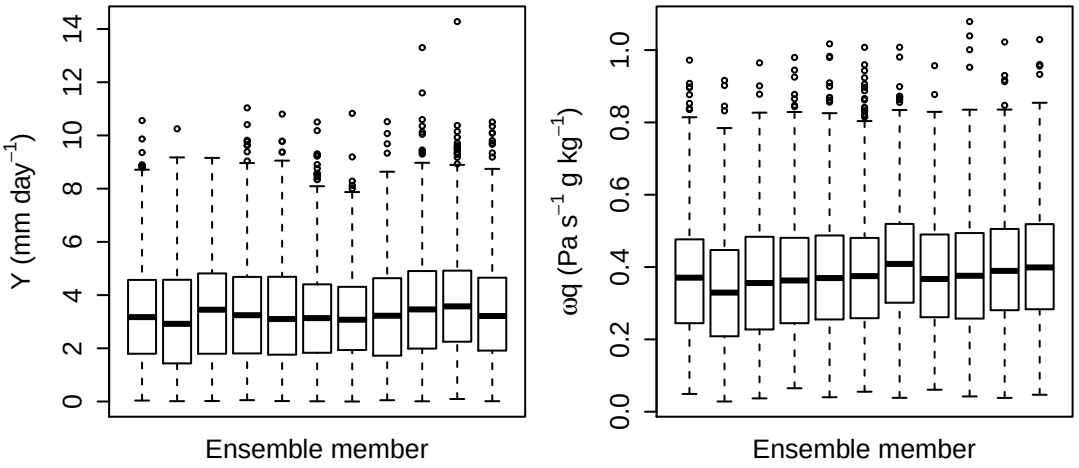

Figure 6.7: Box plots of pentad precipitation (Y) (left panel) and five-day mean moist ascent (ωq) (right panel) averaged over the British Isles, by ensemble member. The control run is on the left of each panel.

to capture the non-linearity in the relationship between moist ascent and the precipitation, i.e. we use a smooth function (f) of the moist ascent predictor to capture its relationship with precipitation in each ensemble member. We use seven parameters for this function, allowing detail in the relationship by placing a knot at the 2nd, 20th, 40th, 60th, 80th and the 98th percentiles of the moist ascent variable in the ensemble. The analysis is not particularly sensitive (not shown) to these placings because the relationship is close to linear anyway. We capture the variation across ensemble members in the relationship by allowing each of the seven parameters that define f to vary with ensemble member via a random effect ($\psi^{(\omega q)}$). Therefore f is a function of ωq and $\psi^{(\omega q)}$. We use another random effect ($\psi^{(0)}$) to allow the intercept of the statistical model to vary with ensemble member.

To capture the symmetrical residual variation in the precipitation suggested by Figure 6.6, we assume a Gaussian residual term, thereby defining our statistical model as a general linear mixed model (LMM) (see Breslow and Clayton, 1993). Note that the Gamma family of distributions is too heavy tailed (not shown) to model the precipitation given the inputs here. Using the Gaussian family, since Figure 6.6 also suggested that the residual variation is smallest at low moist ascent, we allow the residual variation to vary with moist ascent by using a smooth function ($f^{(\sigma^2)}$) of natural cubic spline bases with the same knot placings as with f. The specification of the LMM is, therefore,

$$Y_{ij} \mid \left(\psi_j^{(0)}, \psi_j^{(\omega q)}, \omega q_{ij}\right) \sim \text{Normal}(\mu_{ij}, \sigma_{ij}^2)$$

$$\begin{aligned} \mu_{ij} &= \beta_0 + \psi_j^{(0)} + f(\omega q_{ij}, \psi_j^{(\omega q)}) \\ \log(\sigma_{ij}^2) &= \alpha_0 + f^{(\sigma^2)}(\omega q_{ij}) \end{aligned}$$

$$\left(\psi_j^{(0)}, \psi_j^{(\omega q)}\right) \sim \text{MVN}(\mathbf{0}, \mathbf{\Sigma}) \ ,$$

where i indexes the pentad within each ensemble member, j indexes the member, and μ and σ^2 are the expectation and variance of each fitted Gaussian distribution respectively. We assume that the random effects are multivariate Gaussian with covariance $\mathbf{\Sigma}$.

We use restricted maximum likelihood (REML) (Patterson and Thompson, 1971) to estimate the parameters of the LMM, and display diagnostic residual plots using quantile residuals (see Subsection 4.4.2) in Figure 6.8. Plotting the residuals against fitted values (top-left panel of Figure 6.8), it can be seen that the LMM

can greatly underestimate the precipitation at low fitted values. In contrast the greatest overestimates (by estimated probability) tend to occur at fitted values between 4 and 6 mm day^{-1}, and the lower tail of the quantile residuals is not heavy enough (bottom-right panel). Despite this, overall the fit of the LMM appears to be of sufficient quality. In particular, the LMM seems to represent the precipitation well at high fitted values, which suggests that the extreme precipitation events are represented well. Also see Figure 6.9 for a comparison between the empirical precipitation (top panel) and a sample of precipitation simulated (bottom panel) from the LMM, both plotted against the empirical moist ascent.

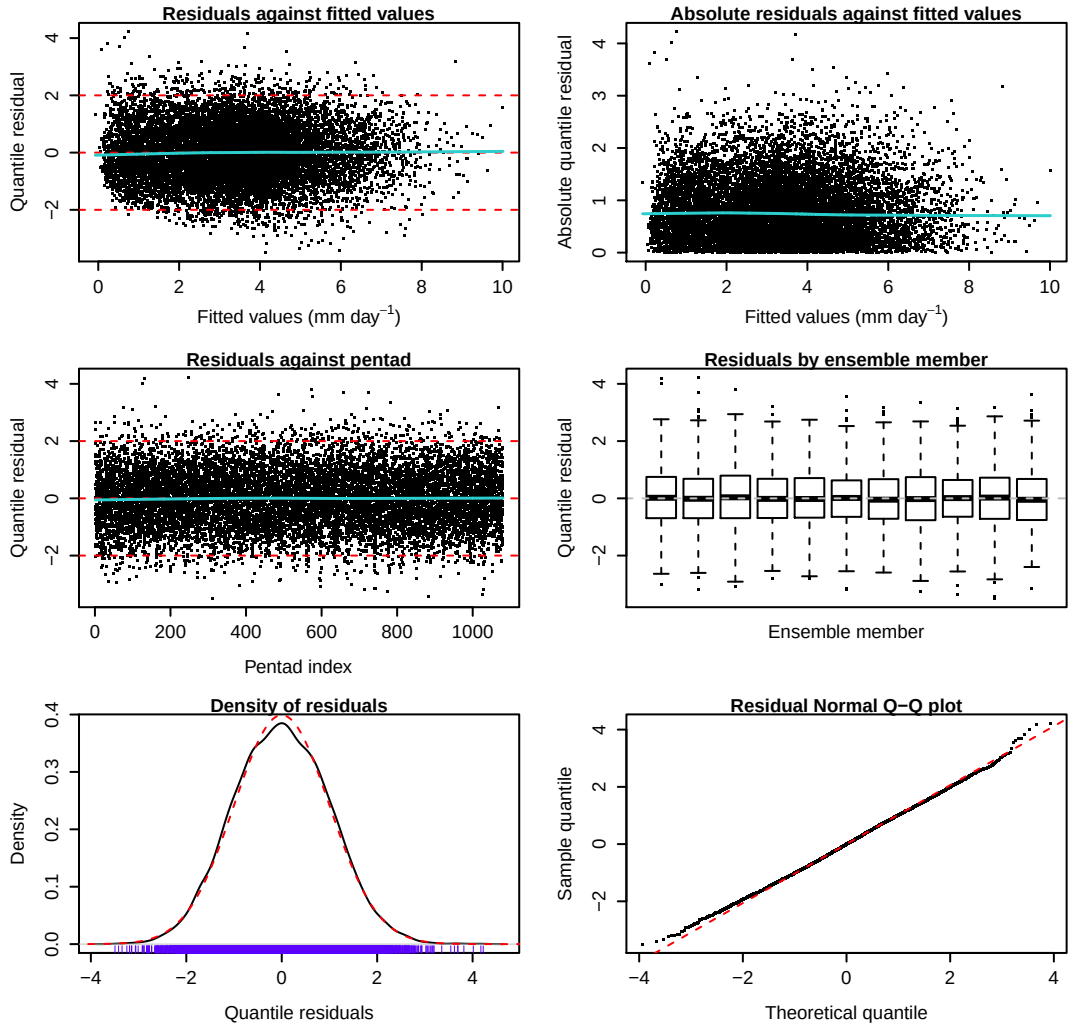

Figure 6.8: Diagnostic residual plots (using quantile residuals) for the LMM. The dashed red lines provide an idealized guide, and the solid blue curves are local polynomial smoothers showing any trends where appropriate.

6.4.4 Decomposition

Figure 6.10 displays the precipitation against moist ascent, conditioned on precipitation greater than 6.8 mm day^{-1} (its 95th percentile). The curves show the

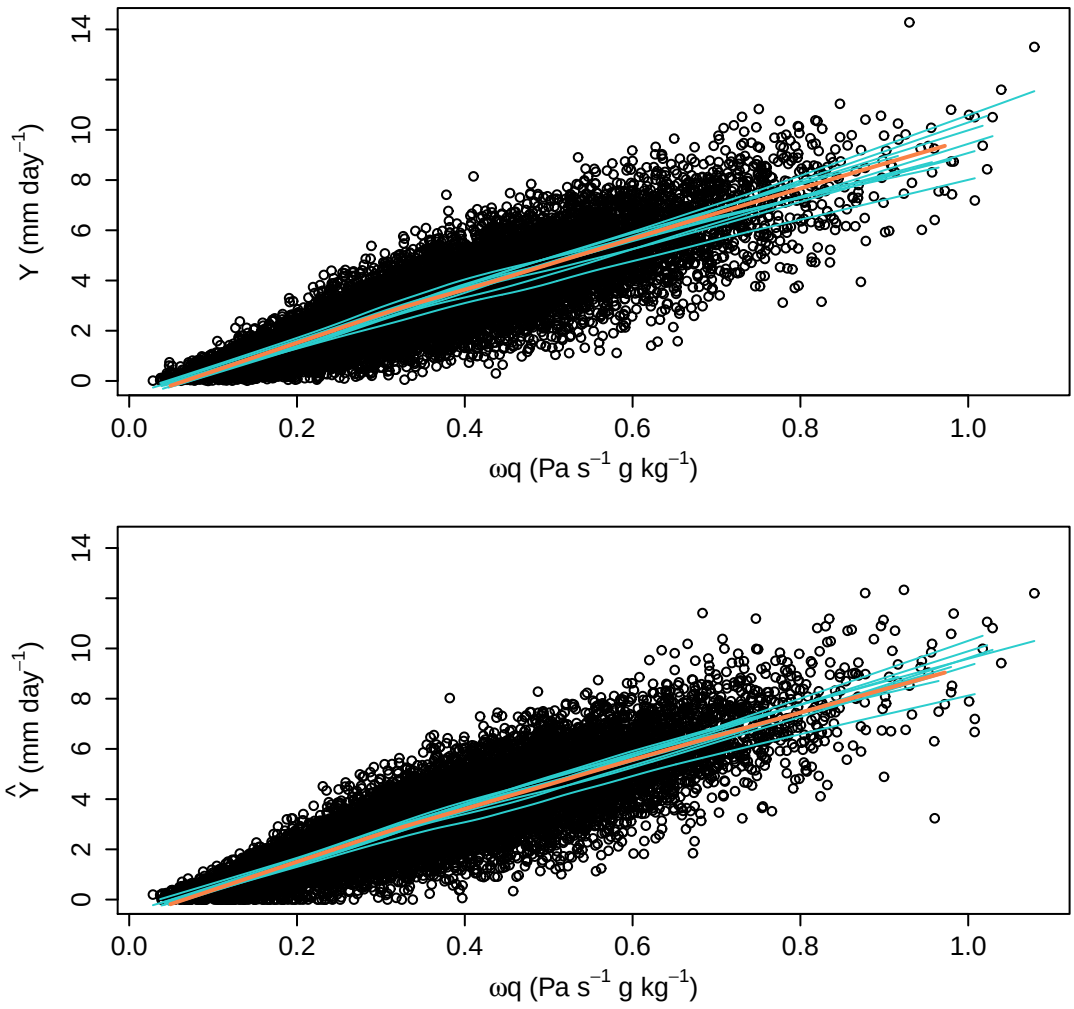

Figure 6.9: LMM diagnostic: Top panel is the response (Y) of the LMM against the predictor (ωq). Bottom panel shows simulated values ($\hat{Y}$) for Y from the LMM. Curves are local polynomial smoothers for each ensemble member, with the control run in orange.

conditional expectation of precipitation estimated by the LMM. As in the unconditional case, there seems little variation across ensemble members in the relationship between moist ascent and precipitation when focusing on the more extreme events. This indicates that winter precipitation extreme events may be simulated robustly across ensemble members. To quantify the extent to which the simulation of these extremes is robust to the uncertainty in the parameters of the RCM, we use the LMM and apply the methodology of Chapter 2 and its extension outlined in this chapter (Section 6.2). We do not restrict the analysis to defining an extreme event as greater than 6.8 mm day^{-1}, however, and instead we repeat the decomposition many times, each with a different threshold up to 13.0 mm day^{-1} (the 99.99th percentile of the empirical precipitation).

Following the methodology of Chapter 2, we separate the predictor (ωq) into its mean (γ) within each ensemble member and its temporal deviations ($\omega q'$) from

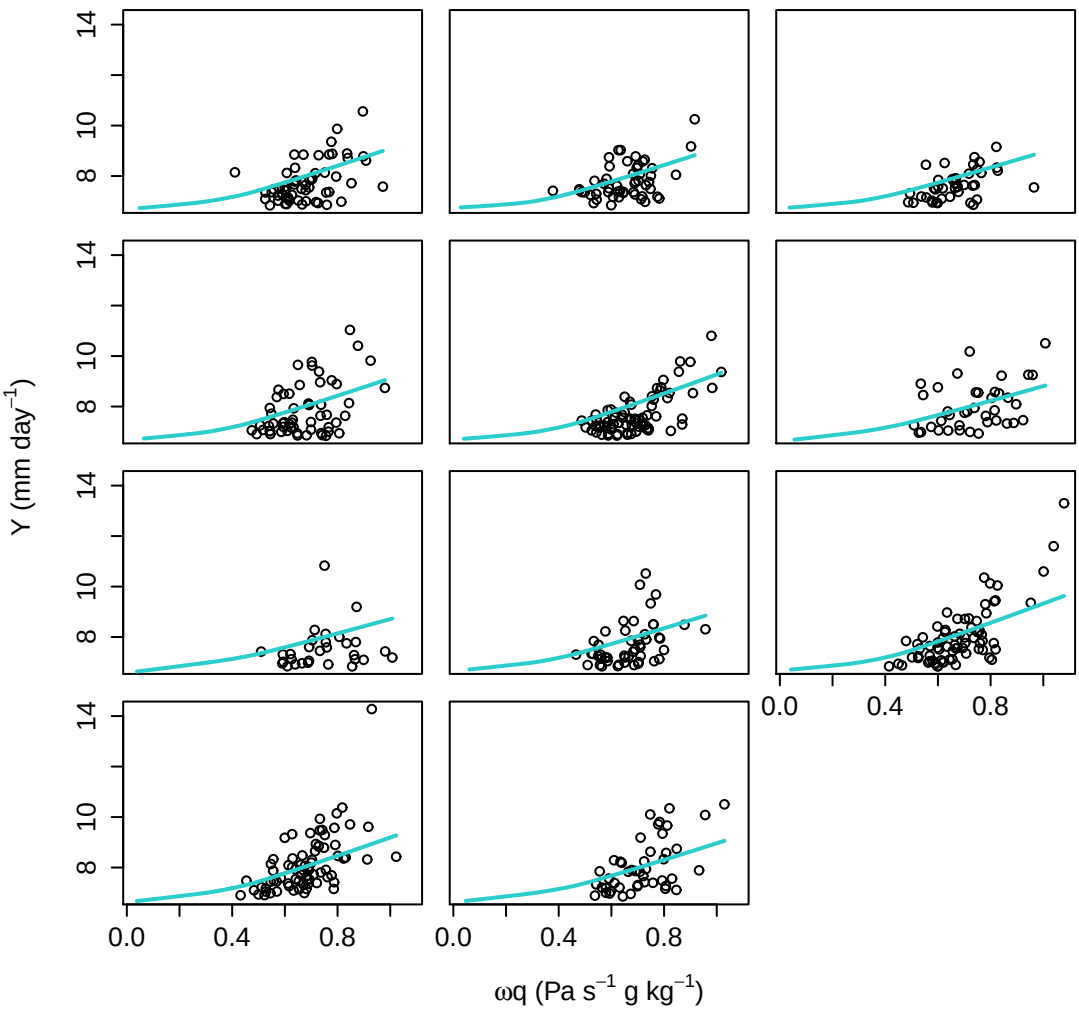

Figure 6.10: Pentad precipitation (Y) against five-day mean moist ascent (ωq) when focusing on precipitation events greater than 6.8 mm day^{-1} in each ensemble member averaged over the British Isles. Curves are local polynomial smoother approximations of the conditional expectation estimated by the LMM given an extreme precipitation event. The control run is top left.

these means, i.e. (including subscripts)

$$\omega q_{ij} = \gamma_j + \omega q'_{ij} \, .$$

Therefore the inputs of the LMM are

- the variation across ensemble members in the intercept, captured by $\psi^{(0)}$,
- the variation across ensemble members in the relationship between moist ascent and the precipitation, $\psi^{(\omega q)}$,
- the variation across ensemble members in the distribution of moist ascent, γ,
- the temporal variation in moist ascent, $\omega q'$,

in addition to the residual term, ϵ.

Following the recommended dependence structure of Subsection 2.2.2 and Section 2.4, we use a multivariate Gaussian copula to model the (unconditional) random effects ($\psi^{(0)}$ and $\psi^{(\omega q)}$) and the predictor means (γ), and sample from the unconditional empirical distribution of the temporal component ($\omega q'$) of the predictor. Therefore we assume independence between the inputs that capture variation due to the differences between the ensemble members and the temporal variation in the unconditional case, although dependence could be present when conditioned on an extreme precipitation event. Assuming independence between $\omega q'$ and the other inputs is supported by the box plots shown by the right panel of Figure 6.7 (page 195), which indicate that the shape of the distribution of ωq is similar across the ensemble members.

Using the model for the inputs described above, and the fitted LMM, we decompose the variation in precipitation using 16 000 samples of size 2000 (i.e. $R = 16\,000$ and $r = 2000$ using the notation of Sections 2.3 and 6.2). We decompose the unconditional variation in the precipitation, before conditioning on precipitation greater than a threshold (t) and using the same samples to decompose the conditional variation in the precipitation, as outlined in Section 6.2 (and Equation (6.1) shown in Subsection 6.3.2). We repeat the conditional decomposition using many different thresholds spanning the distribution of the precipitation, including focusing on only the most extreme events. We do this to ensure that the analysis is not overly sensitive to the choice of threshold, and to examine the extent to which the robustness of the simulation of winter precipitation by the RCM depends on the severity of the precipitation events.

6.4.5 Results

The top panel of Figure 6.11 shows that the empirical (marginal) distribution of winter pentad precipitation over the British Isles is positively skewed, such that the median (not marked) precipitation is 3.2 mm day^{-1} but, for example, the 95th percentile is 6.8 mm day^{-1} and the 99.99th percentile is 13.0 mm day^{-1}. The unconditional and conditional decompositions of the variation in the precipitation is shown in the centre panel of the figure. Different thresholds from zero (i.e. unconditional) up to the 99.99th percentile of the empirical precipitation are used. The decomposition is into 1) the variation solely accounted for by the temporal variation moist ascent, 2) the total variation found to be due to the uncertainty in the RCM parameters, and 3) the remaining unaccounted for (residual) variation, each expressed in absolute terms. The empirical conditional variation in the precipitation is also shown in the centre panel, and is similar to the estimated total conditional

variation in the precipitation (shown as the sum of sources (1) to (3)) at all thresholds less than 9 mm day^{-1}. This indicates that the statistical modelling represents the precipitation well at these thresholds. The empirical conditional variation in the precipitation becomes unreliable at higher thresholds, however, because there are too few data (see bootstrapped confidence intervals in the centre panel, and the rug shown along the horizontal axis of the top panel).

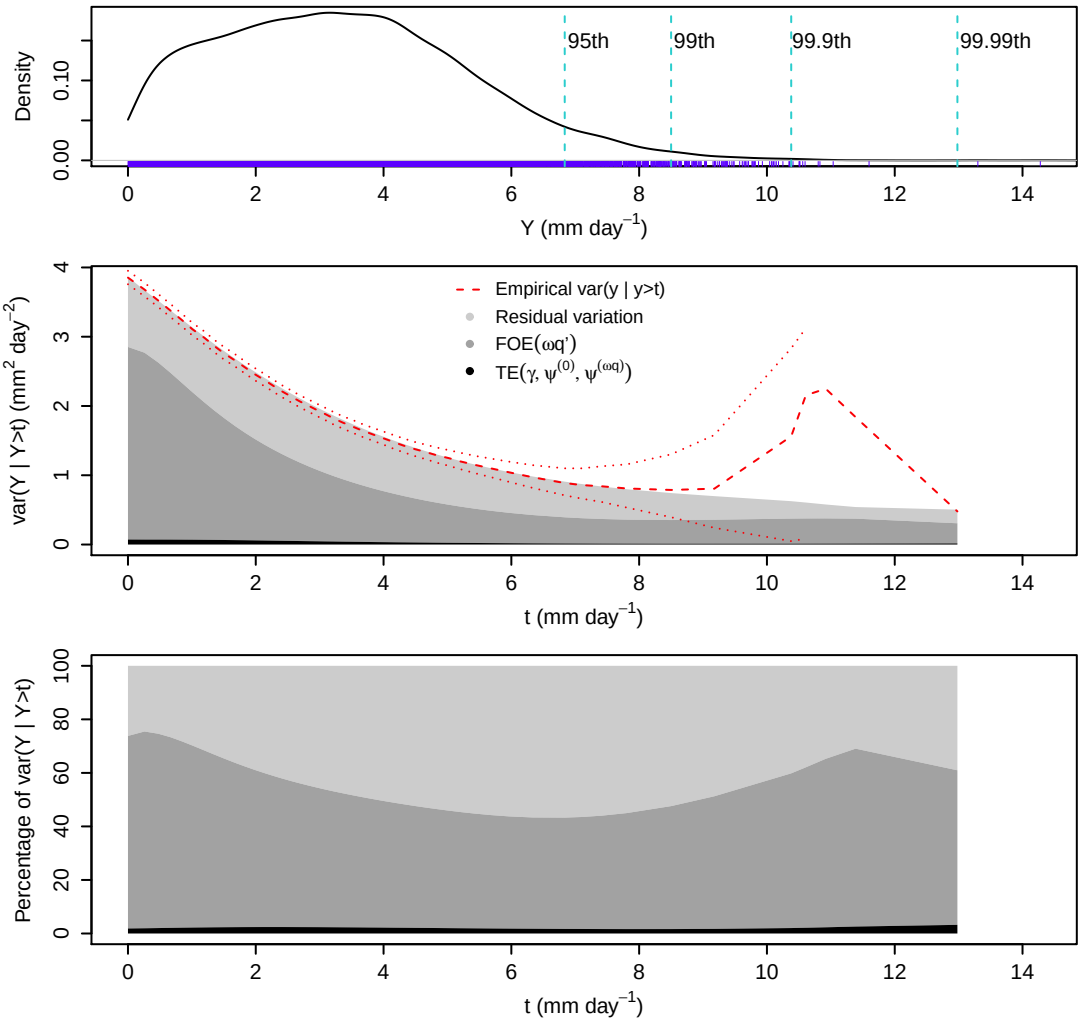

Figure 6.11: Top panel: Kernel density estimate of the empirical winter pentad precipitation (Y) over the British Isles, with some high percentiles marked. Bottom two panels: Conditional decomposition of the estimated total variation (var($Y \mid Y > t$)) in the precipitation into the first-order effect (FOE($\omega q'$)) of the temporal variation in moist ascent, the total effect (TE($\gamma, \psi^{(0)}, \psi^{(\omega q)}$)) of the random effects and the moist ascent means, and the remaining residual variation, for different thresholds (t). These are expressed in absolute terms (centre panel) and as percentages (bottom panel) of var($Y \mid Y > t$). The empirical conditional variation in the precipitation is also shown (dashed red curve) in the centre panel, along with (bootstrapped) 95% confidence intervals.

The centre panel of Figure 6.11 shows that the total conditional variation in the precipitation decreases as the threshold is increased. This is largely a result of the absolute variation in precipitation due to the temporal variation in moist ascent decreasing with threshold, and could be because there is a finite amount of moisture

that the air can hold at a given temperature, and so a finite amount of precipitation that can fall even in extreme conditions. The absolute amount of variation in the precipitation that we find to be due to the uncertainty in the RCM parameters also decreases with threshold, although is small even in the unconditional case.

We assumed independence between the temporal variation in moist ascent and the other inputs in the unconditional case. Conditioning on extreme precipitation events leads to some dependence between the temporal variation in moist ascent and the other inputs, but this dependence is small even using 13.0 mm day^{-1} as the threshold. The greatest correlation between the samples taken for $\omega q'$ and those for one of the other inputs using this high threshold is only -0.12.

The bottom panel of Figure 6.11 shows the decompositions expressed as percentages of the total conditional variation in the precipitation. This panel highlights that the temporal variation in moist ascent accounts for much of the variation in the precipitation, between 42 and 74% depending on the threshold used. This is smallest when the threshold is approximately 6.5 mm day^{-1}, which suggests that more of the variation in light precipitation events and in extreme precipitation events is accounted for by the variation in moist ascent averaged over the British Isles than is the case with moderately heavy precipitation events. This is perhaps because moderately heavy precipitation events may only drive heavy precipitation over a part of the British Isles, whereas the moist ascent is averaged over all of the British Isles. Despite the variation in moist ascent accounting for much of the variation in the precipitation, we find relatively little of the variation (usually less than 3%) in precipitation to be due to the uncertainty in the RCM parameters using any threshold. We therefore find that extreme winter precipitation events over the British Isles are simulated robustly across ensemble members, such that given atmospheric conditions lead to similarly extreme precipitation events regardless of the uncertainty in the values of the RCM parameters.

6.4.6 Discussion

In Chapter 2 we outlined how simulation could be used from a statistical model to decompose the total variation in precipitation in an ensemble of climate models. In this chapter we extended this methodology to focus on extremes. In applying the extended methodology to extreme winter precipitation events over the British Isles in a perturbed physics ensemble, we found that the simulation of these extreme events is robust to the uncertainty in the parameters of HadRM3. We focused on events as extreme as greater than the 99.99th percentile of the empirical precipitation in the ensemble, which are nearly once in 300-winter precipitation events, and found that

the robustness of the precipitation events is not sensitive to the choice of threshold. Instead of using multiple statistical models to do this, we were able to simply weight the simulated samples of the inputs of a single statistical model by their estimated probability of leading to an extreme, and computing the conditional expectation of the precipitation for each sample, for different thresholds. We have, therefore, demonstrated that statistical models do not have to be fitted to a reduced data set to focus on precipitation extremes.

Aside from cloud microphysics, the driving physical processes of large-scale winter precipitation events span multiple grid points of RCMs, and so are represented explicitly by the models instead of being parameterized (recall Section 1.4). As a result, climate models have been found to simulate winter precipitation well generally, and extreme winter events more robustly than extreme summer events (recall Section 1.2 and references therein, e.g. Fowler and Ekström, 2009). Therefore it is perhaps not surprising that we found the representation of winter precipitation events to be largely robust to the uncertainty in the RCM parameters. One might expect, however, that extreme precipitation events might not be simulated as robustly as non-extreme events. Extreme precipitation events are driven by extreme atmospheric conditions, which are rare by their nature and this could exacerbate differences between ensemble members. Despite this, we found that even the most extreme events are robust to uncertainty in RCM parameters, including when focusing only on once in nearly 300-winter precipitation events over the British Isles.

Assuming that HadRM3 simulates extreme winter precipitation events realistically, we suggest that no development is required to use it to analyse such precipitation events. Moreover we suggest that developing HadRM3 by reducing the uncertainty in the values of its parameters would not particularly lead to changes in its representation of extreme large-scale precipitation events. Future work is to repeat our analysis using an ensemble that includes different RCMs. If the simulation of large-scale precipitation extremes is robust across a multi-model ensemble, then the simulation of such extremes can be thought of as robust across RCMs in general, and this would mean that any RCM could be used to analyse the extreme events. Conversely, if substantial disagreement were to be found across multi-model ensembles in the simulation of large-scale precipitation extremes, then decomposing the variation in the precipitation further than in this application (i.e. by decomposing $\mathrm{TE}(\gamma, \psi^{(0)}, \psi^{(\omega q)})$) could help identify the sources of the disagreement. For example there could be disagreement across climate models in what atmospheric conditions lead to extreme events, or in the frequency with which extreme atmospheric conditions are represented. Moreover it could be that the simulation of non-extreme precipitation events is robust across different climate models, but that differences

between the models become apparent when focusing on extreme events. These are examples of the information about climate models that could be gained by using the extended methodology proposed in this chapter.

In addition to the future work suggested above, the sensitivity of the results to the spatial area over which the data are averaged could be examined. For example it could be that extreme events on a smaller spatial scale are simulated less robustly than those averaged over the British Isles. This could be of particular interest because even an event that would be considered extreme (e.g. one of 13 mm day^{-1} averaged over the British Isles) in this application would not necessarily be heavy enough in any given area of the British Isles to cause flooding.

7 Summary, discussion and future work

7.1 Thesis summary

Modern-day general circulation models (GCMs) and regional climate models (RCMs) attempt to represent the climate system of the real world. As such, GCMs and RCMs are tools used to understand the current and past climate, and to make projections of the future climate. Climate models differ, however, because there is considerable uncertainty in how best to represent the climate system of the real world. Although the output of climate models can be compared against observations of the real world to rule out drastically unrealistic models, the differences between climate models remain because comparing their output against observations is limited by the observations available and the variables that are unobserved in the real world, and because climate models may simulate different variables realistically in different atmospheric conditions. An important part of the climate system is precipitation, which is vital for life and can cause drought at one extreme and flooding at the other. Quantifying the extent to which climate models agree in their simulation of variables such as precipitation highlights the extent to which the climate models are affected by the uncertainty that is sampled by the different models. This can give an indication of the reliability of the models in representing the current climate and the clarity in making projections of future climate. Additionally, understanding why the climate models disagree helps their development.

Quantifying the extent to which climate models agree in their simulation of precipitation is, however, not as simple as taking their precipitation output on a given spatial and temporal scale (e.g. daily precipitation totals over Exeter) and comparing the (frequency) distribution from each climate model. Doing this could quantify the extent to which the distribution of precipitation is simulated robustly across the models, but would not quantify the extent to which the simulation of precipitation is robust in any general sense. This is because given atmospheric conditions could lead to different precipitation rates in different climate models, and the atmospheric conditions could be simulated with different frequencies. Such variation across the models in their simulation of precipitation will not necessarily be seen in comparing the distributions of precipitation in each climate model. Moreover, comparing such distributions would not reveal the causes of any disagreement found.

To quantify the extent to which precipitation is simulated robustly across different climate models in a more general sense than restricting the analysis to the distributions of precipitation, and to gain insight into why it varies, we have proposed

analysing the driving physical processes of precipitation. More specifically, we have proposed that the representation of driving physical processes of precipitation by climate models are captured using generalized linear mixed models (GLMMs) and used to estimate the contribution to the variation in the precipitation of the different sources of variation captured. To estimate these contributions we simulate from the GLMM to estimate importance measures, which are tools taken from the mathematical modelling literature. In addition to quantifying the variation in the representation of the physical process that is due to the differences between climate models, importance measures can be used with the GLMM to gain insight into why the precipitation varies both across and within the climate models. This can help understand the temporal variation in the physical process, as well as the cause of the disagreement between models.

In Chapter 3 we used the methodology by relating change in temperature to change in global-mean annual-mean precipitation in response to increased carbon dioxide (CO_2) in an ensemble of GCMs, and decomposed the variation in the precipitation change. In doing this we found that in total a concerningly high proportion (46%) of the variation in the precipitation change in the ensemble is due to the differences between the GCMs. We found that this is largely because the distribution of temperature change varies greatly across different GCMs, and temperature is a main driver of precipitation on this scale. Therefore if the goal is to reduce the sensitivity of the simulated precipitation to the choice of GCM, the GCMs should be developed so as to reduce the sensitivity of the simulated temperature to the choice of GCM. We also found that the total variation in the annual-mean precipitation change that is due to the differences between the GCMs is greater (63%) when averaged over the tropical ocean instead of globally.

By analysing daily-total wet-day summer convective precipitation at the grid-point level over southern Britain in a perturbed physics ensemble of an RCM (HadRM3) in Chapter 4, we found that in total up to 8% of the variation in the convective precipitation is due to the uncertainty in the RCM parameters. We found that this is largely because given atmospheric conditions can lead to different rates of precipitation in different ensemble members, and not because the distributions of precipitation were particularly different in different members. We found the largest driver of convective precipitation to be the variation in atmospheric lapse rate because there is generally enough moisture present in the atmosphere for a strong lapse rate to drive convective precipitation.

In Chapter 5 we used the same perturbed physics ensemble as in Chapter 4 to show how the values of the RCM parameters could be incorporated into a GLMM to

quantify the variation in precipitation that is due to perturbing each individual RCM parameter. Substantial spatial variation was found in the effect on precipitation of perturbing different RCM parameters. We found that the perturbations in the cloud droplet to rain conversion thresholds (CW) and those in the ice-fall speed (VF1) can account for the most variation in summer precipitation at the grid-point level over the British Isles in the ensemble, depending on the area, but that the ensemble is too small to assume that this will generalize to other or larger ensembles.

We extended the methodology to focus on extreme events in Chapter 6, and used this extended methodology to analyse extreme winter pentad (five-day mean) precipitation events averaged over the British Isles in the same perturbed physics ensemble as used in Chapters 4 and 5. We found that the simulation of extreme winter precipitation events is robust to the uncertainty in the RCM parameters, such that given atmospheric conditions are simulated with similar frequencies in each ensemble member and lead to similar precipitation rates. We found that this result is not sensitive to the choice of threshold, above which we defined a precipitation event as extreme. Indeed, the robustness of the events holds using any threshold from analysing all precipitation events to focusing only on once in nearly 300-winter events. We therefore anticipate that developing climate models such that the uncertainty in the parameter values of RCMs is reduced will not particularly change the simulation of large-scale winter precipitation events by the RCMs, including the most extreme events.

7.2 Discussion

The proportion of the variation in precipitation in an ensemble of climate models that is due to the differences between the ensemble members can depend on the spatial and temporal scale on which the precipitation is analysed. For example aggregating the precipitation up from daily-totals to summer-totals (as in Chapter 5) can increase the proportion of variation in the precipitation found to be due to the differences between ensemble members, i.e. increases the signal to noise ratio. This is because there is often great temporal variation on short time scales, and so the impact of the differences between ensemble members on precipitation can be easier to detect in precipitation totals of longer temporal periods, and larger spatial areas.

The disadvantage of aggregating up the temporal and spatial scales of the precipitation, however, is that the direct causal physical link between the represented atmospheric conditions and distinct precipitation events can be lost. Instead, only the physical link between broad atmospheric conditions and the accumulation of precipitation events can be analysed. Precipitation totals on coarse spatial and

temporal scales may be of interest in themselves, but if the goal is to quantify the extent to which the simulation of precipitation is robust across different climate models in a general sense, then perhaps the most informative spatial and temporal scales on which to analyse precipitation are scales that capture distinct precipitation events. This is the reason that we used the finest spatial (the grid-point level) and temporal (daily totals) scales available given the data to analyse convective precipitation in Chapter 4, and the reason that we used pentad precipitation averaged over the land of the British Isles when analysing extreme large-scale precipitation in Chapter 6. Analysing distinct precipitation events allows the causes of the variation across and within climate models in the distinct events to be diagnosed. As a result (for example) in Chapter 4 we found that it is variation across ensemble members in the relationship between the represented atmospheric conditions and the precipitation that causes most of the disagreement between the members in their simulation of convective precipitation events. Furthermore analysing convective precipitation on a local scale allowed us to determine that there is often enough moisture in the atmosphere to mean that the temporal variation in lapse rate is the main driver of convective precipitation events.

Comparing the analyses of Chapters 4 and 6, we find the simulation of large-scale precipitation events to be more robust to the uncertainty in RCM parameters than that of convective precipitation events, including when focusing on only extreme large-scale events. Therefore our results support the notation that convective precipitation is more difficult to represent in climate models than large-scale precipitation because of the fine spatial and temporal scales on which convective precipitation occurs. Whereas we anticipate that the simulation of large-scale precipitation events by RCMs will not particularly change with developments to climate models, their simulation of convective precipitation is set to change as climate models develop.

7.3 Future work

Despite analysing convective precipitation on the finest spatial (the grid-point level of resolution 25 km) and temporal (daily totals) scales available given the data in Chapter 4, convective precipitation generally occurs on even finer scales. Indeed, we were unable to account for much of the variation in convective precipitation with three predictor variables, and this could be because some of their direct physical link with the precipitation was lost because of the resolution of the data. Therefore future work is to repeat the analysis of Chapter 4 using data on a finer spatial and temporal scale to more accurately quantify the robustness of the simulation of convective precipitation events by climate models. In addition to analysing convective precipitation on a finer scale, future work is suggested at the end of each application

chapter. We therefore reserve the rest of this section to suggest possible development to the methodology of Chapter 2.

A possible development to the methodology of Chapter 2 is to include the uncertainty in the fitted parameters of the GLMM when decomposing the variation in precipitation. There will always be uncertainty when using statistical models to estimate the relationships between predictors and the response variable, and these relationships are a part of the physical process of interest. If this uncertainty is included in the analysis then the GLMM could be fitted in a Bayesian framework and the posterior distributions of the parameters used to represent the uncertainty. Indeed, this method is suggested by Gelman and Pardoe (2007) to investigate the size of such uncertainty. Gelman and Pardoe (2007) also suggest attempting to represent the estimated standard errors of the parameters with a multivariate Gaussian distribution when the GLMM is fitted in a frequentist framework.

In Chapters 4 and 5 we fitted a GLMM separately at multiple grid points to decompose the variation in precipitation at each point and examine the spatial pattern. When analysing gridded data, the statistical modelling could be developed by fitting one single spatial statistical model to the data. This would allow the spatial correlations to be captured by the spatial model, and predictors such as a land-sea indicator, longitude, latitude, and height of the land could be used to quantify their influence on precipitation. Moreover the extent to which the effect of such spatial features varies across different climate models could be quantified.

We used two different climate model ensembles for the applications, a multi-model ensemble of GCMs in Chapter 3 and a perturbed physics ensemble of an RCM in the other application chapters. Multi-model ensembles sample the uncertainty in structural equations of climate models, whereas perturbed physics ensembles sample the uncertainty in the values of the physical parameters in a climate model. The relative extent to which the simulation of precipitation is robust to these sources of uncertainty can be quantified by fitting a single GLMM to the precipitation simulated by multiple ensembles. Given a time series from each ensemble member in each ensemble, such a GLMM would require a predictor variable to indicate from which climate model (or which ensemble) each time series came. Indeed, ensembles that sample other sources of uncertainty in climate models can also be included in this way, allowing the uncertainty in initial conditions or in future emission scenarios to be included. Such an analysis could highlight which sources of variation in climate modelling cause the greatest disagreement between models in their representation of important physical processes.

7.4 Concluding remark

The GLMMs outlined in Section 1.9 are a flexible class of statistical models, and are applicable to analysing precipitation in climate model ensembles. As such, this thesis has focused on using importance measures with GLMMs. Despite this, the methodology proposed is potentially applicable to any form of statistical model (e.g. directly to extreme value statistical models, as noted in Chapter 6) fitted to data from an ensemble of climate models, and indeed to applications beyond those in climate science.

References

Allen, M.R. and Ingram, W.J. (2002) Constraints on future changes in climate and the hydrologic cycle. *Nature*, vol 419, pp 224-232

Anderson, B., Borgonovo, E., Galeotti, M. and Roson, R. (2014) Uncertainty in Climate Change Modeling: Can Global Sensitivity Analysis Be of Help? *Risk Analysis*, vol 34, pp 271-293

Asai, T. (1965) A numerical study of the air-mass transformation over the Japan Sea in winter. *Journal of the Meteorological Society of Japan*, vol 43, pp 1-15

Barnett, D.N., Brown, S.J., Murphy, J.M., Sexton, D.M.H. and Webb, M.J. (2006) Quantifying uncertainty in changes in extreme event frequency in response to doubled CO_2 using a large ensemble of GCM simulations. *Climate Dynamics*, vol 26, pp 489-511, doi:10.1007/ s00382-005-0097-1

Bärring, L., Holt, T., Linderson, M.-L., Radziejewski, M., Moriondo, M. and Palutikof, J.P. (2006) Defining dry/wet spells for point observations, observed area averages, and regional climate model gridboxes in Europe. *Climate Research*, vol 31, pp 35-49

Bedford, T. and Cooke, R.M. (2002) Vines − a new graphical model for dependent random variables. *Annals of Statistics*, vol 30, pp 1031-1068

Beniston, M., Stephenson, D.B., Christensen, O.B., Ferro, C.A.T., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylhä, K., Koffi, B., Palutikof, J., Schöll, R., Semmler, T. and Woth, K. (2007) Future extreme events in European climate: an exploration of regional climate model projections. *Journal of Climatic Change*, vol 81, pp 71-95

Birch, C.E., Marsham, J.H., Parker, D.J. and Taylor, C.M. (2014b) The scale-dependence and structure of convergence fields preceding the initiation of deep convection. *Geophysical Research Letters*, vol 41, pp 4769-4776, doi:10.1002/ 2014GL060493

Birch, C.E., Parker, D.J., Marsham, J.H., Copsey, D. and Garcia-Carreras, L. (2014a) A seamless assessment of the role of convection in the water cycle of the West African Monsoon. *Journal of Geophysical Research: Atmospheres*, vol 119, pp 2890-2912, doi:10.1002/ 2013JD020887

Breslow, N.E. and Clayton, D.G. (1993) Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association*, vol 88, pp 9-25

Briffa, K.R., Jones, P.D., Wigley, T.M.L., Pilcher, J.R. and Baillie, M.H.L. (1986) Climate reconstruction from tree rings: Part 2, spatial reconstruction of summer mean sea-level pressure patterns over Great Britain. *Journal of Climatology*, vol 6, pp 1-15

Brooks, S., Gelman, A., Jones, G.L. and Meng, X.-L. (2011) *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC

Buishand, T.A. and Klein Tank, A.M.G. (1996) Regression model for generating time series of daily precipitation amounts for climate change impact studies. *Stochastic Hydrology and Hydraulics*, vol 10, pp 87-106

Byers, H.R. (1965) *Elements of cloud physics*. The University of Chicago Press

Chacón, J.E. and Duong, T. (2010) Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test*, vol 19, pp 375-398, doi:10.1007/ s11749-009-0168-4

Chambers, J.M. (1992) *Linear models*. Statistical Models in S eds, J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole, Chapter 4

Chan, K., Trantanola, S., Saltelli, A. and Sobol', I.M. (2000) *Variance-Based Methods*. Sensitivity Analysis eds, New York: Wiley, pp 167-197

Christensen, J.H., Krishna Kumar, K., Aldrian, E., An, S.-I., Cavalcanti, I.F.A., de Castro, M., Dong, W., Goswami, P., Hall, A., Kanyanga, J.K., Kitoh, A., Kossin, J., Lau, N.-C., Renwick, J., Stephenson, D.B., Xie, S.-P. and Zhou, T. (2013) Climate Phenomena and their Relevance for Future Regional Climate Change. *Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp 1262-1269

Christensen, W.F. and Sain, S.R. (2011) Latent variable modeling for integrating output from multiple climate models. *Mathematical Geosciences*, vol 44, pp 395-410, doi:10.1007/ s11004-011-9321-1

Coe, R. and Stern, R.D. (1982) Fitting models to daily rainfall data. *Journal of Applied Meteorology*, vol 21, pp 1024-1031

Coles, S.G. (2001) *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, Springer-Verlag London Ltd, London

Collins, M. (2007) Ensembles and probabilities: a new era in the prediction of climate change. *Philosophical Transactions of The Royal Society: Series A*, vol 365, pp 1957-1970

Cox, D.R. and Solomon, P.J. (2003) *Components of Variance*. Chapman and Hall/CRC, pp 5

Da Veiga, S., Wahl, F. and Gamboa, F. (2009) Local Polynomial Estimation for Sensitivity Analysis on Models With Correlated Inputs. *Technometrics*, vol 51, doi:10.1198/ TECH.2009.08124

Déqué, M., Rowell, D.P., Lüthi, D., Giorgi, F., Christensen, J.H., Rockel, B., Jacob, D., Kjellström, E., de Castro, M. and van den Hurk, B. (2007) An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Climatic Change*, vol 81, pp 53-70

Della-Marta, P.M., Luterbacher, J., von Weissenfluh, H., Xoplaki, E., Brunet, M. and Wanner, H. (2007) Summer heat waves over western Europe 1880-2003, their relationship to large-scale forcings and predictability. *Climate Dynamics*, vol 29, pp 251-275, doi:10.1007/ s00382-007-0233-1

DeMaris, A. (1993) Odds versus Probabilities in Logit Equations: A Reply to Roncek. *Social Forces*, vol 71, pp 1057-65

Dougherty, R.L., Edelman, A. and Hyman, J.M. (1989) Positivity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation. *Mathematics of Computation*, vol 52, pp 471-494

Dunn, P.K. and Smyth, G.K. (1996) Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, vol 5, pp 236-244

Emanuel, K.A. (1994) *Atmospheric Convection*. Oxford University Press

Easterling, D., Meehl, G., Parmesan, C., Changnon, S., Karl, T. and Mearns, L. (2000) Climate extremes: observations, modeling and impacts. *Science*, vol 289, pp 2068-2074

Fisher, R.A. and Tippett, L.H.C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, vol 24, pp 180-190

Fowler, H.J., Cooley, D., Sain, S.R. and Thurston, M. (2010) Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment. *Journal of Extremes*, vol 13, pp 241-267

Fowler, H.J. and Ekström, M. (2009) Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. *International Journal of Climatology*, vol 29, pp 385-416

Furrer, R., Geinitz, S. and Sain, S.R. (2012) Assessing variance components of General Circulation Model output field. *Environmentrics*, early view, doi:10.1001/ env.2139

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., Rossi, F. and Ulerich, R. (2013) GNU Scientific Library Reference Manual (v1.16), *www.gnu.org/software/gsl/manual/gsl-ref.pdf*

Garratt, J.R. and Physick, W.L. (1983) Low-level wind response to mesoscale pressure scales. *Boundary-Layer Meteorology*, vol 27, pp 69-87

Geinitz, S., Furrer, R. and Sain, S.R. (2012) MMANOVA: A general multilevel framework for multivariate analysis of variance. *Submitted*

Gelman, A. (2005) Analysis of variance: why it is more important than ever. *Annals of Statistics*, vol 33, pp 1-31

Gelman, A. (2008) Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, vol 27, pp 2865-2873, doi:10.1002/ sim.3107

Gelman, A. and Pardoe, I. (2007) Average predictive comparisons for models with nonlinearity, interactions, and variance components. *Sociological Methodology*, vol 37, pp 23-51

Green, P.J. and Silverman, B.W. (1994) *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall, London

Gregory, D. and Rowntree, P.R. (1990) A Mass Flux Convection Scheme with Representation of Cloud Ensemble Characteristics and Stability-Dependent Closure. *American Meteorological Society*, vol 118, pp 1483-1506

Hall, J.W., Sayers, P.B. and Dawson, R.J. (2005) Nation-scale assessment of current and future flood risk in England and Wales. *Natural Hazards*, vol 36, pp 147-164

Hand, W.H., Fox, N.I. and Collier, C.G. (2004) A study of twentieth-century extreme rainfall events in the United Kingdom with implications for forecasting. *Meteorological Applications*, vol 11, pp 15-31

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) *Multivariate Analysis*, London: Academic Press

Hastie, T.J. (1992) *Generalized additive models.* Chapter 7 of Statistical Models in S eds, Wadsworth & Brooks/Cole

Hastie, T.J. and Pregibon, D. (1992) *Generalized linear models.* Statistical Models in S eds, J.M. Chambers and T.J. Hastie, Wadsworth & Brooks/Cole, Chapter 6

Hay, L.E. and Clark, M.P. (2003) Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. *Journal of Hydrology*, vol 282, pp 56-75

Hof, A.F., den Elzen, M.G.J. and van Vuuren, D.P. (2008) Analysing the costs and benefits of climate policy: value judgements and scientific uncertainties. *Global Environmental Change*, vol 18, pp 412-424

Hohenegger, C., Walser, A., Langhans, W. and Schär, C. (2008) Cloud-resolving ensemble simulations of the August 2005 Alpine flood. *Quarterly Journal of the Royal Meteorological Society*, vol 134, pp 889-904

Homma, T. and Saltelli, A. (1996) Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety*, vol 52, pp 1-17

Huebener, H., Cubasch, U., Langematz, U., Spangehl, T., Niehörster, F., Fast, I. and Kunze, M. (2007) Ensemble climate simulations using a fully coupled ocean-

troposphere-stratosphere general circulation model. *Philosophical Transactions of The Royal Society: Series A*, vol 365, pp 2089-2101, doi:10.1098/ rsta.2007.2078

Intergovernmental Panel on Climate Change (2013) Summary for Policymakers in Climate Change 2013: The Physical Science Basis. *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp 20

Jacques, J., Lavergne, C. and Devictor, N. (2006) Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering and System Safety*, vol 91, pp 1126-1134

Jansen, M.J.W. (1999) Analysis of variance designs for model output. *Computer Physics Communications*, vol 117, pp 35-43

Jansen, M.J.W., Rossing, W.A.H. and Daamen, R.A. (1994) *Monte Carlo estimation of uncertainty contributions from several independent multivariate sources*, in J. Gasmanand and G. von Straten (Eds) Predictability and Nonlinear Modelling in National Sciences and Economics. Kluwer Academic Publishers, Donrecht, pp 334-343

Johnson, R.H. and Hamilton, P.J. (1988) The Relationship of Surface Pressure Features to the Precipitation and Airflow Structure of an Intense Midlatitude Squall Line. *Monthly Weather Review*, vol 116, pp 1444-1472

Karl, T.R. and Knight, R.W. (1998) Secular trends of precipitation amount, frequency, and intensity in the United States. *Bulletin of the American Meteorological Society*, vol 79, pp 231-241

Karl, T.R., Knight, R.W. and Plummer, N. (1995) Trends in high-frequency climate variability in the twentieth century. *Nature*, vol 377, pp 217-220

Karl, T., Nicholls N. and Gregory, J. (1997) The Coming Climate. *Scientific American*, vol 276, pp 54-59

Kaufman, C.G. and Sain, S.R. (2010) Bayesian functional ANOVA modeling using Gaussian process prior distributions. *Bayesian Analysis*, vol 5, pp 123-150

Katz, R.W., Parlange, M.B. and Naveau, P. (2002) Statistics of extremes in hydrol-

ogy. *Advances in Water Resources*, vol 25, pp 1287-1304

Kendon, E.J., Rowell, D.P. and Jones, R.G. (2009) Mechanisms and reliability of future projected changes in daily precipitation. *Climate Dynamics*, doi:10.1007/ s00382-009-0639-z

Kendon, E.J., Rowell, D.P., Jones, R.G. and Buonomo, E. (2008) Robustness of future changes in local precipitation extremes. *Journal of Climate*, vol 21, pp 4280-4297

Kessler, E. (1995) On the continuity and distribution of water substance in atmospheric circulations. *Atmospheric Research*, vol 38, pp 109-145

Knight, C.G., Knight, S.H.E., Massey, N., Aina, T., Christensen, C., Frame, D.J., Kettleborough, J.A., Martin, A., Pascoe, S., Sanderson, B., Stainforth, D.A. and Allen, M.R. (2007) Association of parameter, software, and hardware variation with large-scale behaviour across 57 000 climate models. *Proceedings of the National Academy of Sciences of the United States of America*, vol 104, pp 12259-12264

Knutti, R. (2008) Should we believe model predictions of future climate change? *Philosophical Transactions of The Royal Society: Series A*, vol 366, pp 4647-4664, doi:10.1098/ rsta.2008.0169

Knutti, R. (2010) The end of model democracy? *Climatic Change (editorial comment)*, vol 102, pp 395-404

Lambert, F.H., Stine, A.R., Krakauer, N.Y. and Chiang, J.C.H. (2008) How much will precipitation increase with global warming? *Eos, Transactions, American Geophysical Union*, vol 89, doi:10.1029/ 2008EO210001

Lambert, F.H. and Webb, M.J. (2008) Dependency of global mean precipitation on surface temperature. *Geophysical Research Letters*, vol 35, doi:10.1029/ 2008GL034838

Landsman, Z. and Valdez, E.A. (2003) Tail Conditional Expectations for Elliptical Distributions. *North American Actuarial Journal*, vol 7, pp 55-71

Landsman, Z. and Valdez, E.A. (2005) Tail Conditional Expectations for Exponential Dispersion Models. *Astin Bulletin*, vol 35, pp 189-209

Lavers, D.A., Allan, R.P., Wood, E.F., Villarini, G., Brayshaw, D.J. and Wade, A.J. (2011) Winter floods in Britain are connected to atmospheric rivers. *Geophysical Research Letters*, vol 38, doi:10.1029/ 2011GL049783

Lee, Y., Nelder, J.A. and Pawitan, Y. (2006) *Generalized Linear Models with Random Effects: a unified analysis via h-likelihood*. Chapman and Hall/CRC

Lin, Y.-L., Farley, R.D. and Orville, H.D. (1983) Bulk parameterization of the snow field in a cloud model. *Journal of Applied and Meteorology*, vol 22, pp 1065-1092

Little, M.A., Rodda, H.J.E. and McSharry, P.E. (2008) Bayesian objective classification of extreme UK daily rainfall for flood risk applications. *Journal of Hydrology and Earth System Sciences*, vol 5, pp 3033-3060

Marsh, T.J. (2004) The January 2003 flood on the Thames. *Weather*, vol 59, pp 59-62, doi:10.1256/ wea.212.03

Marsham, J.H., Dixon, N., Garcia-Carreras, L., Lister, G.M.S., Parker, D.J., Knippertz, P. and Birch, C.E. (2013) The role of moist convection in the West African monsoon system − Insights from continental-scale convection-permitting simulations. *Geophysical Research Letters*, vol 40, pp 1843-1849, doi:10.1002/ grl.50347

Masson, D. and Knutti, R. (2011) Climate model genealogy. *Geophysical Research Letters*, vol 38, doi:10.1029/ 2011GL046864

McGuffie, K. and Henderson-Sellers, A. (2005) *A Climate Modelling Primer*. 3rd edition, Wiley-Blackwell

McInerney, D.J. and Moyer, E.J. (2012) Direct and disequilibrium effects on precipitation in transient climates. *Atmospheric Chemistry and Physics Discussions*, vol 12, pp 19649-19681, doi:10.5194/ acpd-12-19649-2012

McKay, M.D, Beckman, R.J. and Conover, W.J. (2000) A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code. *Technometrics*, vol 42, pp 55-61

Meehl, G.A., Arblaster, J.M. and Tebaldi, C. (2005) Understanding future patterns of increased precipitation intensity in climate model simulations. *Geophysical Research Letters*, vol 32, L18719

Met Office Hadley Centre (2008) *NCAS British Atmospheric Data Centre: HadRM3-PPE-UK model data.* Available from http://badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dataent_12178667495226008

Min, S.-K., Zhang, X., Zwiers, F.W. and Hegerl, G.C. (2011) Human contribution to more-intense precipitation extremes. *Nature*, vol 470, pp 378-381

Mitchell, J.F.B., Wilson, C.A. and Cunnington, W.M. (1987) On CO_2 climate sensitivity and model dependence of results. *Quarterly Journal of the Royal Meteorological Society*, vol 113, pp 293-322

Montgomery, D.C. (2009) *Design and Analysis of Experiments.* John Wiley & Sons, 7th edition, pp 505-541

Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., van Vuuren, D.P., Carter, T.R., Emori, S., Kainuma, M., Kram, T., Meehl, G.A., Mitchell, J.F.B., Nakicenovic, N., Riahi, K., Smith, S.J., Stouffer, R.J., Thomson, A.M., Weyant, J.P. and Wilbanks, T.J. (2010) The next generation of scenarios for climate change research and assessment. *Nature*, vol 463, pp 747-756

Murphy, J.M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D. and Webb, M. (2007) A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society: Series A*, vol 365, pp 1993-2028, doi:10.1098/ rsta.2007.2077

Murphy, J., Sexton, D., Barnett, D., Jones, G., Webb, M., Collins, M. and Stainforth, D. (2004) Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature*, vol 430, pp 768-772

Nakicenovic, N. and Swart, R. (2000) *Special report on emissions scenarios.* Cambridge, UK: Cambridge University Press

Nelder, J. and Wedderburn, R. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society: Series A*, vol 135, pp 370-384

Nelson, R.B. (2006) *An Introduction to Copulas.* Springer, 2nd ed

Nordhaus, W.D. (2008) *A Question of Balance: Weighing the Options on Global Warming Policies.* New Haven: Yale University Press

Osborn, T. and Maraun, D. (2008) Changing intensity of rainfall over Britain. *Climatic Research Unit information sheet*, no 15

Pall, P., Aina, T., Stone, D.A., Scott, P.A., Nozawa, T., Holberts, A.G.J., Lohmann, D. and Allen, M.R. (2011) Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, vol 470, pp 382-386

Pall, P., Allen, M.R. and Stone, D.A. (2007) Testing the Clausius-Clapeyron constraint on changes in extreme precipitation under CO_2 warming. *Journal of Climate Dynamics*, vol 28, pp 351-363

Palmer, T.N., Doblas-Reyes, F.J., Hagedorn, R. and Weisheimer, A. (2005) Probabilistic prediction of climate using multi-model ensembles: from basics to applications. *Philosophical Transactions of the Royal Society B*, vol 360, pp 1991-1998, doi:10.1098/ rstb.2005.1750

Parker, W.S. (2011) When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science*, vol 78, pp 579-600

Patterson, H.D. and Thompson, R. (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika*, vol 58, pp 545-554

Pope, J.O., Collins, M., Haywood, A.M., Dowsett, H.J., Hunter, S.J., Lunt, D.J., Pickering, S.J. and Pound, M.J. (2011) Quantifying Uncertainty in Model Predictions for the Pliocene (Plio-QUMP): Initial results. *Palaeogeography, Palaeoclimatology, Palaeoecology*, vol 309, pp 128-140

Räisänen, J. (2007) How reliable are climate models? *Tellus*, vol 59, pp 2-29

Ramanathan, V., Crutzen, P.J., Kiehl, J.T. and Rosenfeld, D. (2001) Aerosols, climate and the hydrological cycle, *Science*, vol 294, pp 2119-2124

Ramsay, J. and Silverman, B. (2005) *Functional Data Analysis.* Springer, pp 123,133,136

Roncek, D.W. (1991) Using Logit Coefficients to Obtain the Effects of Independent Variables in Changes in Probabilities. *Social Forces*, vol 70, pp 509-18

Rönnegård, L., Shen, X. and Alam, M. (2010) hglm: A Package for Fitting Hierarchical Generalized Linear Models. *The R Journal*, vol 2, pp 20-28

Sain, S.R. and Furrer, R. (2010) Combining climate model output via model correlations. *Stochastic Environmental Research and Risk Assessment*, vol 24, pp 821-829, doi:10.1007/ s00477-010-0380-5

Sain, S.R., Nychka, D. and Mearns, L. (2011) Functional AVOVA and regional climate experiments: a statistical analysis of dynamic downscaling. *Environmetrics*, vol 22, pp 700-711

Sakamoto, Y., Ishiguro, M. and Kitagawa G. (1986) *Akaike Information Criterion Statistics*. D. Reidel Publishing Company

Saltelli, A. and Annoni, P. (2010) How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, vol 25, pp 1508-1517

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S. (2010) Variance based sensitivity analysis of model output: Design and estimator for the total sensitivity index. *Computer Physics Communications*, vol 181, pp 259-270

Saltelli, A., Chan, K. and Scott, E.M. (2000) *Sensitivity Analysis*. Wiley Series in Probability and Statistics, Chichester: Wiley

Saltelli, A. and Tarantola, S. (2002) On the Relative Importance of Input Factors in Mathematical Models: Safety Assessment for Nuclear Waste Disposal. *Journal of the American Statistical Association*, vol 97, doi:10.1198/ 016214502388618447

Sanderson, B.M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D.J., Ingram, W.J., Piani, C., Stainforth, D.A., Stone, D.A. and Allen, M.R. (2008b) Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. *Journal of Climate*, vol 21, pp 2384-2400

Sanderson, B.M., Piani, C., Ingram, W.J., Stone, D.A. and Allen, M.R. (2008a) Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. *Climate Dynamics*, vol 30, pp 175-190

Sapiano, M.R.P., Stephenson, D.B., Grubb, H.J. and Arkin, P.A. (2006) Diagnosis of variability and trends in a global precipitation dataset using a physically motivated statistical model. *Journal of Climate*, vol 19, pp 4154-4166

Schaller, N., Mahlstein, I., Cermak, J. and Knutti, R. (2011) Analyzing precipitation projections: A comparison of different approaches to climate model evaluation. *Journal of Geophysical Research*, vol 116, doi:10.1029/ 2010JD014963

Sheather, S.J. and Jones, M.C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *Journal of Royal Statistical Society: Series B*, vol 53, pp 683-690

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC

Sobol', I.M. (1976) Uniformly distributed sequences with an addition uniform property. *USSR Computational Mathematics and Mathematical Physics*, vol 16, pp 236-242

Sobol', I.M. (1993) Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, vol 1, pp 407-414

Spencer, P.L. and Stensrud, D.J. (1998) Simulating Flash Flood Events: Importance of the Subgrid Representation of Convection. *Monthly Weather Review*, vol 126, pp 2884-2912

Stainforth, D.A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D.J., Kettleborough, J.A., Knight, S., Martin, A., Murphy, J.M., Piani, C., Sexton, D., Smith, L.A., Spicer, R.A., Thorpe, A.J. and Allen, M.R. (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, vol 433, pp 403-406

Stensrud, D.J. (2007) *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models*. Cambridge University Press

Stephens, G.L., L'Ecuyer, T., Forbes, R., Gettlemen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P. and Haynes, J. (2010) Dreary state of precipitation in global models. *Journal of Geophysical Research*, vol 115, doi:10.1029/ 2010JD014532

Sutton, R.T. and Hodson, D.L.R. (2005) Atlantic Ocean forcing of North American and European summer climate. *Science*, vol 309, pp 115-118

Taylor, K.E., Stouffer, R.J. and Meehl, G.A. (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, vol 93, pp 485-498, doi:10.1175/ BAMS-D-11-00094.1

Tebaldi, C. and Knutti, R. (2007) The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of The Royal Society: Series A*, vol 365, pp 2053-2075

Trenberth, K.E., Dai, A., Rasmussen, R.M. and Parsons, D.B. (2003) The changing character of precipitation. *American Meteorological Society*, forum, pp 1205-1217

Vidale, P.L., Lüthi, D., Frei, C., Seneviratne, S.I. and Schär, C. (2003) Predictability and uncertainty in a regional climate model. *Journal of Geophysical Research*, vol 108, doi:10.1029/ 2002JD002810

Wallace, J.M. and Hobbs, P.V. (2006) *Atmospheric Physics: An Introductory Survey.* Second edition, Academic Press Publications

Wand, M.P. (2000) A comparison of regression spline smoothing procedures. *Computational Statistics*, vol 15, pp 443-462

Weiss, N.A. (2005) *A Course in Probability.* Addison-Wesley, pp 385-386

Wilby, R.L. and Wigley, T.M.L (1997) Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, vol 21, pp 530-548, doi:10.1177/ 030913339702100403

Wilson, P.S. and Toumi, R. (2005) A fundamental probability distribution for heavy rainfall. *Geophysical research letters*, vol 32, L14812

Wood, S.N. (2006) *Generalized Additive Models: An Introduction with R.* Chapman and Hall/CRC Press

Xu, C. and Gertner, G.Z. (2007) Extending a global sensitivity analysis technique to models with correlated parameters. *Comput Stat Data Anal*, doi:10.1016/ j.csda.2007.04.003

Yee, T.W. and Stephenson, A.G. (2007) Vector generalized linear and additive extreme value models. *Extremes*, vol 10, pp 1-19

Yip, S., Ferro, C.A.T. and Stephenson, D.B. (2011) A simple, coherent framework for partitioning uncertainty in climate projections. *Journal of Climate*, vol 24, pp 4634-4643, doi:10.1175/ 2011JCLI4085.1